

MASTEROPPGAVE
Master i yrkespedagogikk
Mai 2020

Vurdering av muntlige eksamensprestasjoner i høyere utdanning

En studie av interbedømmer reliabilitet ved eksamen i Etterforskning på Politihøgskolen



Av: Christine Sætre Nordsletten

OSLOMET

OsloMet – storbyuniversitetet

Fakultet for lærerutdanning og internasjonale studier

Institutt for yrkesfaglærerutdanning

Forord

Masteroppgaven utgjør siste del av et fireårig deltidsstudium ved OsloMet. Gjennom studiet har jeg gjennomført totalt fire prosjekter ved min arbeidsplass Politihøgskolen (PHS), med utgangspunkt i selvvalgte problemstillinger og ved bruk av forskningsmetode. Jeg har fått anledning til å fordype meg i tema som teknologistøttet læring og vurdering av studentprestasjoner, og jeg har samlet data fra både studenter og fagansatte. Tusen takk til alle dere som har deltatt i prosjektene. Studiet har vært lærerikt og krevende. Personlig har jeg utviklet min kompetanse faglig, pedagogisk og forskningsmessig, og jeg har lært mye om meg selv. Særlig den pedagogiske kunnskapen har medført konstruktive diskusjoner i egen faggruppe, og bidratt til utvikling av nye læringsaktiviteter for våre studenter – med formål å best mulig tilrettelegge for læring. Forhåpentligvis har studiet også vært nyttig for PHS gjennom de ulike prosjektene og tilhørende kunnskapsutvikling.

Valg av tema for masteroppgaven bunner i en interesse og nysgjerrighet som er utviklet over tid. Jeg har lenge lurt på hvordan vurdering foregår på tvers av sensorer og kommisjoner. Gjennom prosjektet håper jeg å bidra til økt kunnskap om interbedømmer reliabilitet og hva som danner grunnlag for karakterbedømminger. Forhåpentligvis vil mine funn kunne brukes som beslutningsgrunnlag i det videre arbeidet med vurdering av studentprestasjoner.

Vi har hatt et flott og dedikert lærerteam. Tusen takk til Arne Roar Lier og Evi Schmid som har ledet studiet på en veldig god måte, og til Hæge Nore, min fantastiske veileder gjennom tre år, som alltid er positiv og på tilbudssiden. Dine konstruktive tilbakemeldinger har gitt meg mye å tenke og gruble på, og har vært av stor betydning for min læring. Takk også til Leif Langli, som var veileder første året. Tusen takk til Elinor Hunnestad, som har vært min studiekollega, venninne, kritiker, motivator, støtte og veileder gjennom disse årene. Og tusen takk til Nina Sunde for gjennomlesning og nyttige tilbakemeldinger.

Da jeg startet studiet var barna mine, Marie og Ole Kristian, 13 og 11 år. Nå er de 17 og 15. Det setter ting i perspektiv, og viser at jeg har brukt mange og viktige år av livet på dette arbeidet. Tusen takk for at dere har vært så tålmodige og forståelsesfulle. Tusen takk til min kjære samboer, Rune Christian Johnsen, som har vært en uvurderlig støtte og diskusjonspartner, og til hans sønn Eirik, som tålmodig har funnet seg i at kjøkkenbordet har vært overfylt av bøker og PC. Tusen takk til mamma, pappa og lillebror Arild, som alltid heier på meg og alltid er støttende og omsorgsfull.

Sammendrag

I dette prosjektet undersøker jeg hvordan interne sensorer i emnet Etterforskning ved Politihøgskolen (PHS) vurderer muntlige eksamensprestasjoner. Emnet utgjør 18 studiepoeng og består av fem fag. Selv om det er et lovfestet krav at studentenes læringsutbytte skal prøves på en rettferdig og upartisk måte som sikrer utdanningskvaliteten, har vi lite kunnskap om hvordan muntlige eksamensprestasjoner vurderes på tvers av sensorer ved PHS. Det finnes også lite forskning om temaet, både nasjonalt og internasjonalt. Karakterer brukes i utstrakt grad i høyere utdanning, og fungerer som seleksjonsfiltre ved ansettelse. De kan ha stor betydning for den enkeltes studie- og jobbmuligheter, og for arbeidsgiveres mulighet til å sikre seg ønskede kandidater. Kunnskap om hvordan vi vurderer er dermed nødvendig for å kunne fatte kvalifiserte slutninger om utdanningskvaliteten.

Våren 2018 fikk jeg samtykke fra to studenter til å filme deres muntlige eksamensprestasjoner. Opptakene ble vist til 29 interne sensorer, som – ved å besvare spørsmål i et digitalt spørreskjema – vurderte og karaktersatte prestasjonene. I forlengelsen av de individuelle vurderingene gjennomførte jeg fire fokusgruppeintervjuer, hvert bestående av tre sensorer med ulik fagtilknytning, som til sammen utgjorde en eksamenskommisjon. I intervjuene ble kommisjonsmedlemmene bedt om å enes om en felles karakter, slik de ville gjort ved en reell eksamensgjennomføring.

Funnene viser at sensorene vurderer ulikt – i karakterer, begrunnelser, spørsmålsstilling og i bruken av kriterier. Mens noen har en holistisk tilnærming, er andre mer analytisk orientert. Mange baserer vurderingen på sitt profesjonelle skjønn og på implisitte kriterier, heller enn på eksplisitte kriterier. Karakterene varierer i spennet A til C/D for oppgave 1, og C/D til F for oppgave 2. Spennet er nesten like stort innad i samme fag som på tvers av fagene.

Vurderingene gjøres som hovedregel med utgangspunkt i det faget sensor representerer, og kommisjonskarakteren synes å bli satt et sted midt mellom kommisjonsmedlemmenes individuelle vurderinger. Forhold ved kommunikasjonen – både i argumentasjon, innhold og rekkefølge – synes å ha betydning i samspillet når karakter skal settes. Begrunnelsene som karakteren bygger på kjennetegnes tidvis av å være lite utdypet. Dermed kan det være vanskelig å forstå hva som konkret var styrker og svakheter ved prestasjonen.

Funnene gir grunn til å rette større oppmerksomhet mot vurderingsfeltet for å sikre rettferdig, gyldig og pålitelig bedømming av studentenes kompetanser.

A study of inter-rater reliability of oral examinations at the Norwegian Police University College (PHS)

In this study, I examine the inter-rater reliability of oral exam performances in the course *Investigation* at PHS. The course consists of five subjects. Although there is a statutory requirement that students' learning outcomes need to be rated in a fair and impartial manner, we have limited knowledge on how oral exam performance is evaluated across assessors at PHS. There is also modest research on the topic, both nationally and internationally. Grades are widely used within higher education, often as selection filters in hiring processes. They can be of great importance to students' future job opportunities and for employers' possibilities to hire the most desired candidates. Knowledge on how student performance is assessed, is necessary in order to make qualified conclusions about the quality of education.

In the spring of 2018, two students agreed to have their oral exams videotaped, for the purpose of my project. I showed the recordings to 29 internal assessors at PHS, who each rated the performance by answering questions in a digital questionnaire. The questions not only addressed the grading of the performance, but also the premises that formed the basis for the grading. I also conducted four focus group interviews, each consisting of three assessors with different professional affiliation. Each focus group constituted an examination board. In the interviews, the assessors were asked to come to an agreement about the grading of the performance, as they would in the context of a real exam.

My findings show that the assessors evaluate differently – in grades, premises, questioning and in the use of criteria. While some have a holistic approach, others are more analytically oriented. The assessments are based largely on implicit criteria and the assessors' professional judgments, rather than on explicit criteria, and on the assessors' professional affiliation. The grades vary from A to C/D for task 1, and from C/D to F for task 2. In the examination boards, the grade is set somewhere in between the board members' individual assessments. Aspects of communication – in argumentation, content and order of the arguments – seem to be of importance when the grade is to be determined in the boards. The premises on which the grades are based are sometimes characterized by being inadequately reasoned. Thus, it can be difficult to fully understand the strengths and weaknesses of the performance.

In the future, we should pay greater attention to the area of assessment to ensure fair, valid and reliable assessments of student competencies.

Innholdsfortegnelse

FORORD	1
SAMMENDRAG	2
A STUDY OF INTER-RATER RELIABILITY OF ORAL EXAMINATIONS AT THE NORWEGIAN POLICE UNIVERSITY COLLEGE (PHS)	3
INNHALDSFORTEGNELSE	4
OVERSIKTER OVER FIGURER	5
1. INNLEDNING	6
1.1 BAKGRUNN FOR VALG AV OPPGAVE	6
1.2 PROBLEMSTILLING, FORSKNINGSSPØRSMÅL OG AVGRENSNING	8
1.3 HVORFOR ER PROSJEKTET VIKTIG?	9
1.4 MITT KUNNSKAPSSYN	12
1.5 OPPBYGGING OG STRUKTUR	13
2. EKSAMEN OG VURDERING	14
2.1 EKSAMEN I EMNET ETTERFORSKNING	14
2.2 KORT HISTORISK TILBAKEBLIKK	16
2.3 LOVVERK OG SENTRALE DOKUMENTER	17
2.4 DEFINISJON OG BEGREPSFORKLARINGER	18
2.5 FORSKNING PÅ FELTET	23
3. TEORETISKE PERSPEKTIVER	24
3.1 AUTENTISK VURDERING	24
3.1.1 <i>Fem dimensjoner for autentisk vurdering</i>	25
3.2 VALIDITET OG RELIABILITET I VURDERING OG EKSAMEN	27
3.3 RETORIKK OG BRUK AV OVERTALELSESMIDLER I VURDERING	30
3.3.1 <i>Overtalelsesmidlene (appellformene)</i>	32
3.4 SUBJEKTIVITET OG SKJØNNSUTØVELSE I VURDERING	33
4. METODE	36
4.1 MIN FORFORSTÅELSE	37
4.2 PÅLITELIGHET, GYLDIGHET OG TRANSPARENS	38
4.3 METODISK TILNÆRMING	39
4.3.1 <i>Videoopptak av eksamensprestasjoner</i>	39
4.3.2 <i>Spørreskjema og fokusgruppeintervju</i>	40
4.4 UTVALG AV INFORMANTER	45
4.5 DATAINNSAMLING	46
4.6 TRANSKRIBERING	48
4.7 ETISKE BETRAKTNINGER	50
5. ANALYSE OG FUNN	51
5.1 ANALYSEPROSESSEN	52
5.2 FUNN	56
5.2.1 <i>Oppsummering av hovedfunn</i>	57
5.2.2 <i>Vurdererne vurderer ulikt – i karakterer, begrunnelser og spørsmålsstilling</i>	58
5.2.3 <i>Kriterier og standarder brukes og oppfattes ulikt</i>	62
5.2.4 <i>Ulike tilnærminger i argumentasjon og overtalelse</i>	66
5.3 REFLEKSJONER RUNDT FUNNENES GENERALISERBARHET	70
6. DRØFTING	71

7. OPPSUMMERING OG VEIEN VIDERE	81
LITTERATURLISTE	85
VEDLEGG	90

Oversikter over figurer

FIGUR 1: KARAKTERBEDØMMING VED HENHOLDSVIS OPPRINNELIG SENSUR (NEDERSTE LINJE) OG KLAGESENSUR (SØYLER) FOR TOTALT 250 KLAGESAKER. FOR EKSEMPEL SER MAN AT STUDENTER SOM KLAGET PÅ EN C, BLE VURDERT I SPENNET MELLOM A OG F VED KLAGESENSUR.	11
FIGUR 2: ENDELIG KARAKTERBEDØMMING I DE TILFELLER HVOR OPPRINNELIG SENSUR OG KLAGESENSUR AVVEK MED TO KARAKTERER ELLER MER. NEDERSTE RAD VISER OPPRINNELIG KARAKTER, DERETTER VISES KARAKTER GITT VED KLAGESENSUR, OG TIL SLUTT, I SØYLENE, ENDELIG KARAKTER.	12
FIGUR 3: KODESYSTEM ETTER NEDBRYTING TRINN 2 – EMPIRISK ORIENTERT.....	54
FIGUR 4: KODESYSTEM ETTER NEDBRYTING TRINN 3 – TEORETISK ORIENTERT	55
FIGUR 5: INDIVIDUELLE KARAKTERBEDØMMINGER MED TILHØRENDE FAGTILKNYTNING.....	58
FIGUR 6: KOMMISJONSKARAKTERER FOR HVER FOKUSGRUPPE OG FOR DEN OPPRINNELIGE KOMMISJONEN SOM VURDERTE KANDIDAT 1, SAMT KOMMISJONSMEDLEMMENES INDIVIDUELLE KARAKTERBEDØMMINGER	59

1. Innledning

Temaet for denne oppgaven er vurderernes vurderinger. Den handler om hvordan interne sensorer ved Politihøgskolen (PHS) vurderer muntlige eksamensprestasjoner i emnet Etterforskning. Eksamenen gjennomføres mot slutten av siste studieår i bachelorutdanningen, og er den mest omfangsrike i studiet med sine 18 studiepoeng. Den foregår ved at kandidatene redegjør for en helhetlig etterforskning av et beskrevet scenario (eksamensoppgaven), for en kommisjon bestående av tre sensorer. Sensorene representerer tre av de fem fagene som til sammen utgjør etterforskningsemnet. Rundt 7000 studenter har gjennomført emnet i løpet av de siste ti årene. Mange sensorer, både interne og eksterne, er involvert i vurderingsarbeidet, som strekker seg over to uker. I denne oppgaven ønsker jeg å få mer kunnskap om hvordan muntlig eksamen i etterforskningsemnet vurderes, og hva som vektlegges ved karaktersettingen – individuelt og i samspillet mellom sensorer. Prosjektet kan bidra til å kaste lys over kvaliteten på vurderingsarbeidet ved PHS, og til å gjøre ordningen mer transparent. Den kan også bidra til å skape godt beslutningsgrunnlag for det videre vurderingsarbeidet ved høgskolen generelt og innen etterforskningsemnet spesielt. Kanskje kan mine funn også være til nytte for andre i utdanningssektoren.

1.1 Bakgrunn for valg av oppgave

Eksamen utgjør en sentral del av vurderingssystemet i høyere utdanning, og består i dag av skalaene bestått/ikke bestått og A-F. De er fastsatt i Lov om universiteter og høyskoler (UH-loven), og ble innført med Kvalitetsreformen for å få en felles nasjonal karakterskala for all høyere utdanning, i tråd med internasjonale systemer (St.meld. nr. 27 (2000-2001), s. 32-33; Universitets- og høgskolerådet, 2008, s. 2). Karakterene skal gjenspeile i hvilken grad en student har oppnådd de læringsutbytter som er satt som mål for emnet eller studiet (Kunnskapsdepartementet, 2014, s. 39). En karakter representerer dermed en verdi langs en skala som er ment å gi informasjon om en students kompetanse, og som kan brukes av utdanningsinstitusjoner og potensielle arbeidsgivere for å rangere studenter. En sentral årsak til innføringen av et felles nasjonalt karaktersystem, var nettopp ønsket om å gjøre kvalifikasjonene mer forståelige og sammenlignbare – for studenter, utdanningsinstitusjoner, arbeidslivet og allmennheten. Det skulle gjøres ved å klargjøre den enkeltes kunnskap, ferdigheter og generelle kompetanse (Kunnskapsdepartementet, 2014, s. 8-9).

Regjeringen har klare forventninger til at utdanningsinstitusjonene organiserer

vurderingsarbeidet på en måte som inngir tillit hos studenter, utdanningsinstitusjoner og potensielle arbeidsgivere. Sensurordningene må være egnet til å prøve studentenes læringsutbytte på en rettferdig og upartisk måte som sikrer utdanningskvaliteten (Meld. St. 16 (2016-2017); NOU 2020:3; UH-loven, 2005). Ved PHS er det et overordnet mål å tilby utdanninger av høy kvalitet som er relevante for politiet og samfunnet. Utdanningene skal bidra til at politiet kan utføre sine oppgaver på en måte som ivaretar borgernes rettssikkerhet og samfunnets interesser (Politihøgskolen, 2017). For å ivareta disse ambisiøse målsetningene, er det vesentlig at studentenes karakterer gjenspeiler den kompetansen de er ment å representere. Det fordrer at vurdering foregår innenfor et felles rammeverk hvor samme prestasjon vurderes (tilnærmet) likt på tvers av sensorer.

Mitt valg av tema og problemstilling er begrunnet i, og motivert av, forhold på både individ- og samfunnsnivå. På individnivå veier rettferdighetshensyn og studentenes rettsikkerhet tungt. Utdanningsinstitusjoner og arbeidsgivere har ofte ikke kjennskap til studentene utover dokumentasjonen som vedlegges ved studie- eller jobbsøknader, og karakterene blir dermed viktige indikatorer for å skille mellom flinke og mindre flinke studenter (Strøm, Falch, Gunnes & Haraldsvik, 2013, s. 9). Dermed kan karakterene få stor betydning for den enkeltes muligheter for videre studier og jobb. Eksamen i etterforskningsemnet er omfattende og krevende. For noen studenter ligger det hundrevis av arbeidstimer til grunn for eksamensprestasjonen, for andre betydelig mindre. Generelt bør det være en sammenheng mellom studieinnsats og resultater. Imidlertid viser undersøkelser at det er store variasjoner i forholdet mellom karakternivå og tidsbruk på studiearbeid på tvers av utdanninger (Meld. St. 16 (2016-2017), s. 20; Strøm et al., 2013).

Sensurordninger og karaktersetting har også betydning for potensielle arbeidsgivere og samfunnet. Gjennom flere år har det vært mangel på jobber til nyutdannet politi, og konkurransen om utlyste stillinger er hard. Etterforskning er et fagfelt som innebærer inngripen i enkeltpersoners skjebner (for eksempel i form av pågripelser og fengslinger), og feil kan få store konsekvenser både for enkeltindivider og for rettssikkerheten.

Politidistriktene må kunne ha tillit til at karakterbedømmingen gjenspeiler kandidatens kompetanse i lys av de fastsatte læringsutbytter, slik at de kan gjøre riktige prioriteringer ved ansettelse og sikre seg de mest kompetente personene ut fra sitt behov. I lys av ovennevnte vil kvaliteten i vurderingsarbeidet også kunne ha betydning for tilliten til både politiet og politiutdanningen.

Jeg underviser selv innen etterforskningsemnet og har deltatt som sensor en rekke ganger. Min erfaring er at vurderinger gjøres ulikt – til dels svært ulikt – på tvers av kommisjoner og individer. Imidlertid får jeg, som sensor, kun innblikk i et knippe av kommisjonene. Ved PHS har vi lite kunnskap om hvordan vi vurderer på tvers av sensorer og kommisjoner, og om hvorvidt eksamensform og vurdering gir et korrekt bilde av studentenes læringsutbytte. Tross omfattende bruk av karakterer i høyere utdanning, finnes det lite forskning – nasjonalt og internasjonalt – på hva som ligger til grunn for de vurderinger som gjøres ved karaktersetting, og hvordan vurderingene gjøres på tvers av sensorer. Ovennevnte har ført meg frem til min problemstilling.

1.2 Problemstilling, forskningsspørsmål og avgrensning

I dette prosjektet vil jeg undersøke problemstillingen:

Hvordan vurderer interne sensorer ved Politihøgskolen muntlige eksamensprestasjoner?

Jeg ønsker å få mer kunnskap om hva som legges til grunn for karaktersettingen på tvers av sensorer, og hva som skjer i samspillet mellom sensorer når de skal enes om en felles karakter. Derfor har jeg valgt følgende forskningsspørsmål:

1. Hvilke begrunnelser legges til grunn for karakterbedømmingene?
2. Hvordan vektlegges de ulike vurderingene i samspillet når karakter skal settes?

Det er ikke min intensjon å undersøke kvaliteten på de faglige begrunnelsene. Jeg ønsker å undersøke problemstillingen i et mer overordnet perspektiv. Finnes det trekk, eller karakteristika, ved begrunnelsene som kan belyse hvordan vi vurderer på tvers av sensorer, og hvordan vurderingene vektlegges i samspillet når karakter skal settes? Det interessante er sammenstillingen og vurderingene totalt sett, og ikke den enkeltes vurderinger. For eksempel kan det dreie seg om hvordan de utarbeidede kriterier og standarder brukes ved karakterbedømmingen. Eller om sensorene synes å ha de samme fokus i vurderingene. Ved å ha et mer overordnet fokus håper jeg å kunne bidra til funn som er generaliserbare også til andre kontekster og aktører.

Avgrensning

Problemstillingen vil bli undersøkt med utgangspunkt i muntlig eksamen i emnet Etterforskning ved PHS. Ved gjennomføring av denne eksamen benyttes både interne og eksterne sensorer. I denne oppgaven har jeg avgrenset til kun å se på interne sensorers vurderinger. Med det mener jeg personer som er oppført på skolens sensorliste i etterforskningsemnet, og som hadde et aktivt ansettelsesforhold til PHS på tidspunkt for datainnsamlingen. Grunnen til avgrensningen var at jeg antok at det kunne bli utfordrende å få eksterne sensorer til å sette av den nødvendige tid som deltakelse ville ha krevd (cirka 2-3 timer). En annen årsak var at jeg ønsket å utvikle kunnskap om hvordan fagansatte innen samme institusjon (PHS) vurderer samme prestasjon.

Mine forskningsdeltakere blir i oppgaven betegnet som *informanter*. Jeg bruker også begrepene *vurderer* og *sensor* om disse. Om *empiri* bruker jeg også benevnelsen *innsamlet data*. Kvale og Brinkmann presiserer at data i en kvalitativ studie ikke er samlet inn, men utviklet i fellesskap i samspillet mellom forsker og forskningsdeltakere (Kvale & Brinkmann, 2015, s. 218). Når jeg bruker innsamlet data mener jeg de data som har fremkommet gjennom datainnsamlingen.

1.3 Hvorfor er prosjektet viktig?

Samfunnsdebatten er stadig preget av diskusjoner rundt eksamen og karaktersetning. Et søk på internett gir mangfoldige treff på artikler og debattinnlegg om tilfeldigheter ved karaktersetning, med overskrifter som *Sensurlotto* og *Student klaget og gikk fra B til F* (se for eksempel Hagen, 2016; Mathy, 2018; Mellingsæter, Torset & Husøy, 2018; Ørjasæter, 2016). I et debattinnlegg i Khrono i april 2020 stiller to professorer spørsmål om hvorfor vi trenger eksamen og hvorfor det er nødvendig å teste hva noen kan på et spesifikt tidspunkt. Blant annet mener de at det er en «kortslutning» å tro at vurderingskriterier sørger for like rammevilkår for alle, fordi disse alltid må ses i sammenheng med eksamenskonteksten og forhold som dagsform, eksamenslokaler og sensorenes kompetanse og erfaring (Burner & Berge, 2020). Artikler i *Politiforum* har over tid rettet søkelyset mot de dårlige jobbutiktene for ferdigutdannet politi, og tidvis også på karakterenes betydning for jobbmuligheter (se for eksempel Inderhaug, 2019). I disse dager er vurderingsordningene i høyere utdanning oppe til debatt (NOU 2020:3). I forslag til ny UH-lov pekes det på utfordringer knyttet til

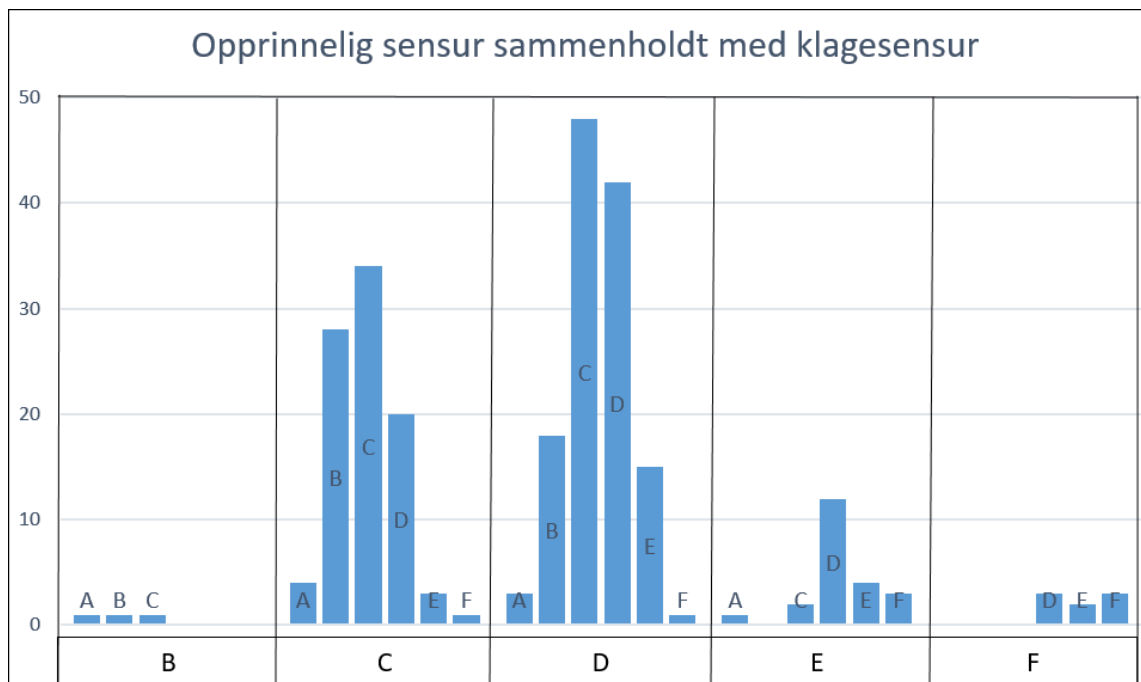
vurderingssamsvar. Blant annet diskuteres gjeninnføring av to sensorer, krav til filming av muntlig eksamen og innføring av en tretrinns karakterskala av typen utmerket/bestått/ikke bestått – som tiltak for å sikre utdanningskvaliteten i form av en uhildet og rettferdig karaktersetting (s. 203-204).

I 2016 ble det iverksatt et sensurprosjekt ved PHS hvor formålet var å heve kvaliteten på sensurarbeidet (Politihøgskolen, 2016b). Gruppen gjennomførte en spørreundersøkelse blant et utvalg fagansatte, og gjennomgikk alle dokumenter som omhandlet eksamen og sensur. De fant at de mange ulike fagtradisjonene som inngår i politiutdanningen, kan utgjøre en utfordring i vurderingssammenheng: «Sammenlignet med studier ved andre høyskoler og universitet er nok potensialet for bred og mangfoldig subjektivitet vel så stort ved Politihøgskolen» (s. 92). Derfor anbefalte gruppen å igangsette et forskningsprosjekt for å «fremskaffe systematisk kunnskap om den subjektiviteten som ligger til grunn for sensur blant interne sensorer ved PHS» (s. 92). Det ble ikke gjort.

Ved PHS formidles resultater av vurderingsarbeidet gjennom en årlig kvalitetsrapport. Kvalitetsrapporten som var gjeldende for tidspunkt for min datainnsamling, synes å rapportere om god utdanningskvalitet (Politihøgskolen, 2019a). Der fastslås at etterforskningseksamen er godt tilpasset læringsutbyttet for emnet, men at det knyttes noe usikkerhet til hvorvidt studentene får nok ferdighetstrening i å gjennomføre og evaluere etterforskningsmetoder. Videre fremgår at eksamenen har blitt godt kvalitetssikret gjennom flere år og at emnet er faglig relevant for politigeneralisten. Hva som ligger til grunn for disse slutningene, fremgår i liten grad. For eksempel er emnets faglige relevans for politigeneralisten kun rapportert gjennom ordet «ja», som er studieleders svar på spørsmålet om relevans. Øvrige studieledere har rapportert tilsvarende for sine hovedområder. Av rapporten fremgår at eksamensformen vil bli endret (fra muntlig til skriftlig) med ny programplan, fordi den er administrativt krevende å gjennomføre. Blant annet er det utfordrende å få tak i nok sensorer.

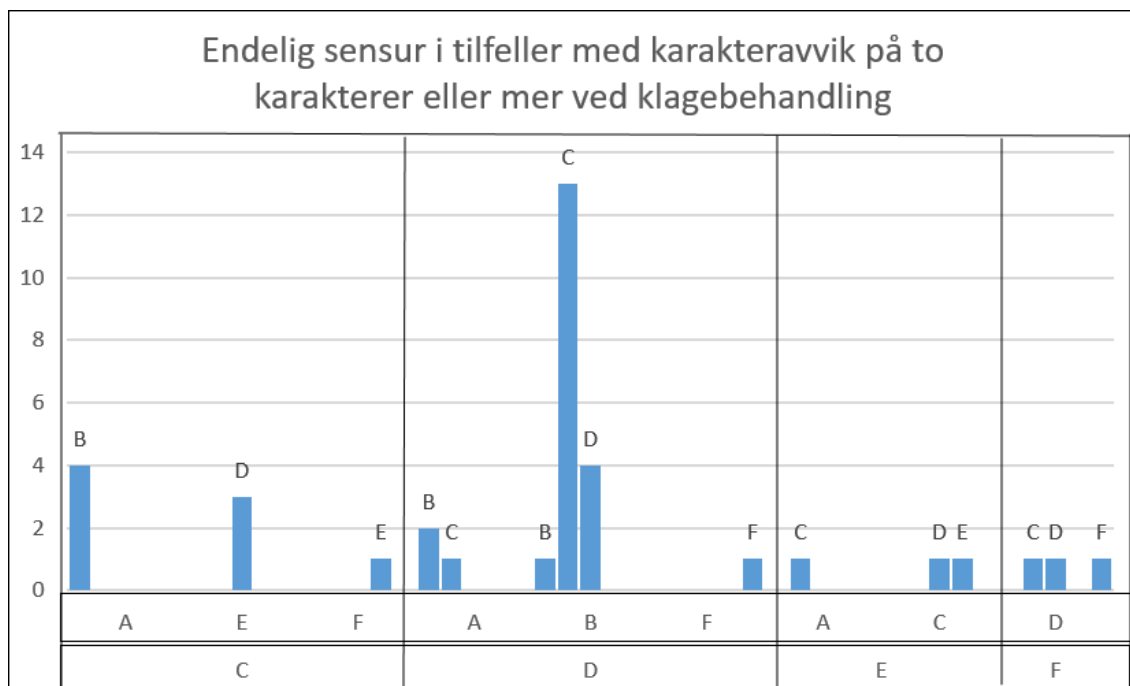
PHS deltok nylig i en undersøkelse av Framtida.no, hvor man så på hvor mange karakterklager utdanningsinstitusjoner innen høyere utdanning har mottatt og behandlet i tidsrommet 2009-2018, totalt 130.000 klager (Aadland, 2019). Tallene viser at kun 51 % av klagen ved PHS (alle studiesteder) endte med samme karakter som ved opprinnelig sensur. Jeg har selv hentet inn tall fra sensur og klagesensur for skriftlige eksamener gjennomført i tre emner ved PHS høsten 2018, totalt 250 eksamensbesvarelser. Dataene ble oversendt

(anonymisert) fra skolens studieavdeling. Min sammenstilling og analyse av tallene viser store forskjeller i sensorenes vurderinger. Kun 84 av de 250 klagen endte med samme karakter ved første og andre sensur. I 122 tilfeller gikk klagen til gunst for studenten og i 44 tilfeller til ugunst. Studentprestasjoner som ved opprinnelig sensur ble vurdert til C, D eller E, ble ved klagesensur bedømt i hele karakterspennet mellom A og F:



Figur 1: Karakterbedømming ved henholdsvis opprinnelig sensur (nederste linje) og klagesensur (søyler) for totalt 250 klagesaker. For eksempel ser man at studenter som klaget på en C, ble vurdert i spennet mellom A og F ved klagesensur.

I 36 av de 250 klagesakene (14,4 %) var karakterspriket på minst to karakterer. I fire tilfeller var karakteravviket på tre karakterer, mens det i ett tilfelle var på hele fire karakterer. I tråd med UH-lovens § 5-3 (6) ble det foretatt en ytterligere vurdering i disse tilfellene, før endelig karakter ble satt. I figur 2 vises endelig karakterbedømming for alle 36 saker, samt karakter gitt ved henholdsvis opprinnelig sensur og klagesensur. Det kan synes som om endelig karakter som hovedregel ble plassert et sted midt imellom de to første vurderingene. For eksempel ble C og A ble til B, C og E til D, og C og F til E.



Figur 2: Endelig karakterbedømming i de tilfeller hvor opprinnelig sensur og klagesensur avvok med to karakterer eller mer. Nederste rad viser opprinnelig karakter, deretter vises karakter gitt ved klagesensur, og til slutt, i søylene, endelig karakter.

Gjennomgangen viser lav interbedømmer reliabilitet i vurderingene, og er i seg selv vektige grunner til at vi trenger mer kunnskap om hvordan vi vurderer.

1.4 Mitt kunnskapssyn

Kvalitative forskere vil alltid ha med seg et sett av antagelser om verden, et kunnskapssyn, som styrer eller rettleder deres forskning (Postholm, 2010, s. 33). Dette studiet bygger på et sosialkonstruktivistisk kunnskapssyn. Sosialkonstruktivisme, slik det er beskrevet av Berger og Luckmann (1966), vektlegger at virkeligheten er samfunnsskapt og må forstås i lys av sosiale faktorer (Tjora, 2017, s. 27). Sosialkonstruktivisme tar utgangspunkt i epistemologi (læren om hva kunnskap er). Innenfor dette paradigmet ses kunnskap som konstruksjon av mening og forståelse som skapes i møtet mellom mennesker som samhandler. Kunnskap er ikke noe som er gitt og som skal tilegnes, men noe som er i stadig endring og utvikling (Postholm, 2010, s. 21). Det faktum at ulike mennesker kan ha vidt forskjellige oppfatninger av samme fenomen, kan forklares godt innenfor det konstruktivistiske paradigmet (Tjora, 2017, s. 27). Den sosiokulturelle retningen innenfor sosialkonstruktivismen ser læring som noe kollektivt som skjer i samspillet mellom individer og omgivelsene, i kombinasjonen mellom ytre samspillsprosesser og indre bearbeidingsprosesser (Postholm, 2010, s. 23-25). Språket og det kommunikative er det viktigste verktøyet for kunnskapsutvikling og forståelse,

fordi det er gjennom kommunikasjon at individet blir delaktig i kunnskaper og ferdigheter (Säljö, 2001, s. 38). All kommunikasjon forutsetter en kontekst for å bli begripelig, og våre handlinger vil alltid inngå i kontekster (fysiske, sosiale, kommunikative, kulturelle, historiske). Det innebærer at konteksten noe ytres eller handles i, er en del av det som gir ytringen en bestemt betydning (Postholm, 2010, s. 22; Säljö, 2001, s. 138-139).

I dette studiet ønsker jeg å belyse hvordan vurdering foregår på tvers av sensorer, og i samspillet når karakter skal settes. I et historisk og kulturelt perspektiv er karaktersystemet menneskeskapt og sosialt konstruert, og et resultat av økonomiske, sosiale og politiske prosesser (se kapittel 2.2). Det har som formål å gi informasjon om kompetanse på en måte som er tilnærmet lik og sammenlignbar på tvers av institusjoner og nasjoner. Denne konteksten utgjør et viktig bakteppe for problemstillingen. Mitt kunnskapssyn har påvirket mine valg i prosjektet på mange måter. For eksempel er innsamlet data i stor grad basert på åpne spørsmål, fordi jeg ønsket å belyse problemstillingen med utgangspunkt i informantenes forståelser og oppfatninger. Mine refleksjoner rundt egen forforståelse (kapittel 4.1) og dennes betydning for forskningen, springer ut fra erkjennelsen av at jeg som forsker påvirker kunnskapen som utvikles – i alle prosjektets faser. Min fortolkning av dataene gjennom analysen er et resultat av at jeg anser kunnskap som noe som utvikles og konstrueres, og ikke noe som er gitt. Valg av teori rundt autentisitet henger sammen med mitt kunnskapssyn om at mennesker lettere kan reprodusere og anvende kunnskap og ferdigheter dersom konteksten ligner på den konteksten kunnskapen og ferdighetene ble lært i. Til slutt i prosjektet oppsummerer jeg heller enn å konkludere, fordi jeg erkjenner at andre undersøkelser gjennomført i andre kontekster og med andre forskningsdeltakere, kan komme til andre resultater.

1.5 Oppbygging og struktur

I kapittel 2 vil jeg beskrive den kontekstuelle rammen for prosjektet, gi et kort historisk tilbakeblikk for dagens karaktersystem, redegjøre for overordnede rammer og styringsdokumenter, definere sentrale begreper for problemstillingen og redegjøre for tidligere forskning på feltet. I kapittel 3 redegjør jeg for fire teoretiske perspektiver som problemstillingen og mine funn kan ses i lys av. Først ser jeg på en teoretisk modell for autentisk vurdering utviklet av nederlandske forskere, som beskriver fem dimensjoner for autentisk vurdering. Til rammeverket og kompetansebasert vurdering hører også reliabilitet og

validitet, som er viktige kvalitetsindikatorer i alt vurderingsarbeid. De blir behandlet i et eget delkapittel. Deretter vil jeg redegjøre for noen retoriske perspektiver som jeg anser interessante i lys av forskningsspørsmål nummer 2, herunder hvilke overtalelsesmidler, eller appellformer, som benyttes i samspillet når karakter skal settes. Denne tilnærmingen ble mer og mer aktuell etter hvert som jeg bearbeidet og analyserte dataene, fordi jeg gjennom disse prosessene oppdaget ulike forhold ved kommunikasjonen og argumentasjonen som syntes å ha betydning for karakterbedømmingen i kommisjonene. Til slutt i teorikapittelet ser jeg på skjønnsutøvelse i vurdering og hvordan subjektiviteten kan påvirke hva vi ser, vektlegger og verdsetter. I kapittel 4 gir jeg en detaljert redegjørelse for mine metodiske valg og overveielser, herunder hvordan dataene er samlet inn og hvilke spørsmål som er brukt i datainnsamlingen. Jeg beskriver også mulige svakheter ved prosjektet som jeg har reflektert over underveis, og jeg redegjør for etiske betraktninger. Kapittel 5 innledes med en beskrivelse av hvordan jeg har gått frem for å gjøre meg i stand til å analysere dataene (analyseprosessen). Denne delen kunne vært plassert under kapittel 4, men jeg anser den som så tett knyttet til funnene at det var naturlig å plassere analyseprosess og funn i samme kapittel. Videre i kapittel 5 følger beskrivelse av hovedfunnene, etterfulgt av en mer detaljert utdypning av disse. Til slutt gjør jeg noen refleksjoner omkring funnenes generaliserbarhet. I kapittel 6 drøfter jeg funnene i lys av teori, tidligere forskning og problemstillingen. Kapittel 7 oppsummerer prosjektet, herunder funnenes betydning for PHS. Til slutt knytter jeg noen refleksjoner til veien videre.

2. Eksamen og vurdering

I det følgende vil jeg beskrive hvordan muntlig eksamen i emnet Etterforskning er innrettet og gjennomføres. Deretter gir jeg en kortfattet oversikt over de historiske hovedlinjene som har ført frem til dagens karaktersystem, og over lovverk og sentrale dokumenter som regulerer eksamen i høyere utdanning. Videre definerer og operasjonaliserer jeg sentrale begrep for prosjektet, og redegjør til slutt for forskning innen temaet, nasjonalt og internasjonalt.

2.1 Eksamen i emnet etterforskning

Etterforskningsemnet gjennomføres i tredje studieår av bachelorutdanningen, og utgjør 18 studiepoeng, det vil si 10 % av den treårige bachelorutdanningens totale antall studiepoeng. Emnet består av fem fag: strafferett- og prosess (juss), rapport- og etterforskningslære (rel),

psykologi (psy), kriminalteknikk (kri) og digitalt politiarbeid (dipo). Eksamen er muntlig, og gjennomføres i løpet av en 2-ukers periode i juni. I Oslo er mellom 40 og 50 studenter (to klasser) oppe til eksamen hver dag, fordelt på ti kommisjoner.¹ I tillegg gjennomføres tilsvarende eksamen ved PHS' lokasjoner i Bodø og Stavern i samme periode. Kommisjonene består av tre sensorer, interne og eksterne, som representerer hver sine fag. De interne er hentet fra bachelorutdanningens første og tredje studieår, samt fra avdeling for etter- og videreutdanning. Det er noe rulling av interne sensorer på tvers av lokasjoner, men som hovedregel sensurerer man ved egen lokasjon. Kommisjonssammensetningene endres fra dag til dag. Noen kommisjoner består bare av interne sensorer, noen av en ekstern og to interne og (unntaksvis) to eksterne og en intern. Noen sensorer er med i hele perioden, mens andre kun deltar en dag eller to. Fagene juss og rel er representert i alle kommisjoner, i tillegg til ett av de tre øvrige fag. Prestasjonen skal vurderes i lys av alle fem fag.

I oppgaven bes studentene om å redegjøre for en helhetlig og formålsstyrt etterforskning som ivaretar krav til kvalitet, effektivitet, rettssikkerhet og forebygging av justisfeil – med utgangspunkt i en praktisk case de får utdelt en time før de skal inn til kommisjonen. Til samme tid får de vite hvilken fagsammensetning kommisjonen har. Den påfølgende timen brukes til individuelle forberedelser i et klasserom med eksamensvakter, før studenten føres til kommisjonen. Studenten får en halvtime til å presentere sin etterforskning, etterfulgt av et kvarter med spørsmål fra sensorene. Sensur gis på slutten av hver eksamensdag ved at kandidaten blir kalt inn til kommisjonen for å få sin karakter, og – hvis ønskelig – en begrunnelse. Studentene kan ikke klage på karakteren, kun på formelle feil.

I forkant av eksamensgjennomføringen ble det avholdt fire sensormøter, hvorav to i Oslo, ett i Stavern og ett i Bodø. Hver sensor var forventet å delta på ett av møtene. Det er utarbeidet relativt omfattende (eksplisitte) vurderingskriterier for emnet i form av et 13 siders skriv. I tillegg er det for hver eksamensoppgave utarbeidet fem sensorveiledninger, en for hvert fag. På sensormøtet ble det fra studieleder uttrykt en forventning om at sensorene setter seg inn i sensorveiledningene for de oppgaver de skal sensurere, særlig for de to fagene som ikke er representert i kommisjonen.

¹ Det gjaldt for tidspunkt da jeg gjennomførte min datainnsamling. Fra 2019 ble det tatt inn færre studenter til politiutdanningen, og tallet vil bli ytterligere redusert fra 2020. Det får virkning for tredje studieår to år senere.

2.2 Kort historisk tilbakeblikk

Forut for 1999 opererte man med en rekke ulike karakterskalaer i Norge. Mange av dem var fingraderte, og ble brukt forskjellig på tvers av institusjoner og fagområder (Det norske universitetsråd, 2000, s. 57). Norges deltakelse i Bologna-prosessen fra 1999 markerer et tidsskille for høyere utdanning (Gynnild, 2010, s. 1). Arbeidet førte til endringer i det norske og europeiske utdanningssystemet, i gradsstrukturer, studieplaner og karaktersystemer. I 1999 nedsatte Universitetsrådet en arbeidsgruppe som skulle utarbeide forslag til nytt nasjonalt karaktersystem for høyere utdanning, i tråd med europeiske karaktersystemer (Universitets- og høyskolerådet, 2008). Arbeidsgruppens forslag var å innføre de to skalaene bestått/ikke bestått og A-F (Det norske universitetsråd, 2000). Karaktertrinnene A-F tilsvarer trinnene i det europeiske European Credit Transfer System (ECTS), som er et system for å kunne sammenligne karakterer på tvers av land og institusjoner. Forslaget ble tiltrådt av departementet i Stortingets kvalitetsmelding for høyere utdanning (St.meld. nr. 27 (2000-2001), s. 32-33). I juli 2002 ble dagens karaktersystem tatt inn i Lov om universiteter og høyskoler, etter forutgående tilråding fra Utdannings- og forskningsdepartementet (2002). I løpet av høsten 2003 ble karaktersystemet innført i hele UH-sektoren. Som oppfølging kom Utdannings- og forskningsdepartementet, i brev av 10.5.2004, med nærmere retningslinjer for bruk av det nasjonale karaktersystemet. Der fastslås at det er den verbale beskrivelsen av prestasjoner på hvert karakternivå, som skal danne utgangspunkt for karaktersettingen (Utdannings- og forskningsdepartementet, 2004).² Universitets- og høyskolerådet (UHR) har i ettertid utviklet generelle, kvalitative beskrivelser for hvert karaktertrinn i samsvar med departementets retningslinjer (Universitets- og høyskolerådet, 2011). Som en videreføring av internasjonale prosesser, ble Nasjonalt kvalifikasjonsrammeverk for høyere utdanning fastsatt av Kunnskapsdepartementet i mars 2009. Med det kom også krav om at studentenes kvalifikasjoner, i alle studieprogram, skulle beskrives gjennom læringsutbytter. I 2011 ble kvalifikasjonsrammeverket for høyere utdanning innlemmet i Nasjonalt kvalifikasjonsrammeverk for livslang læring (NKR), som sist ble oppdatert i 2014 (Kunnskapsdepartementet, 2014).

De siste tiårene har vi beveget oss fra en testkultur i retning av en vurderingskultur (Baartman, Bastiaens, Kirschner & van Der Vleuten, 2007, s. 116). De to kulturene kan ses som hver sine ytterpunkter av en skala. Mens testkulturen fokuserer på å teste grunnleggende

² Gjentatt i brev 14. desember 2011 (Meld. St. 16 (2016-2017), 2017, s. 56)

ferdigheter og kognitive kompetanser, fokuserer vurderingskulturen på utvikling meningsfulle og autentiske læringsaktiviteter som er relevante for arbeidslivet (Baartman et al., 2007, s. 116). I utviklingen har det vokst frem en rekke nye vurderingsmetoder som i litteraturen er omhandlet med mange navn, herunder *kompetansebasert vurdering* og *autentisk vurdering* (Baartman et al., 2007, s. 117). Jeg vil komme tilbake til disse begrepene.

2.3 Lovverk og sentrale dokumenter

Det norske karaktersystemet er fastsatt i Lov om universiteter og høyskoler (UH-loven, 2005). Lovens § 3-9 sier at testing og vurdering av kandidatenes kunnskaper og ferdigheter må foregå på en faglig betryggende og upartisk måte, og at vurderingen skal sikre det faglige nivået i studiet. I samme paragraf lovfestes krav til skriftlig sensorveiledning som skal være egnet til å sikre en så ensartet vurdering som mulig. Systemet er kriteriebasert, og vurdering av eksamen skal gjøres med utgangspunkt i hvorvidt, og i hvilken grad, en student har nådd de læringsutbytter som er satt som mål for emnet eller studieprogrammet. Samtidig skal hver student, uavhengig av utdanningsnivå, vurderes opp mot hele karakterskalaen, med utgangspunkt i de verbale beskrivelsene av prestasjoner på hvert nivå (Meld. St. 16 (2016-2017), s. 56; Utdannings- og forskningsdepartementet, 2004).

Studietilsynsforskriften (2017) lovhjemler at studietilbudet skal være faglig oppdatert og relevant for arbeidslivet, og at vurderingsformene skal være tilpasset læringsutbyttene (§ 2-2). Også Kunnskapsdepartementet uttrykker en klar forventning om at utdanningene må være relevante. Studentene skal forberedes på arbeidshverdagen de vil møte etter endt utdanning, på en måte som gjør dem i stand til å bruke sin kompetanse i arbeidslivet (Meld. St. 16 (2016-2017), s. 16). Det er i tråd med *Forskrift om studier og eksamen ved Politihøgskolen* (2017), hvor en hovedmålsetning er å «sikre utdanning av reflekterte og handlekraftige polititjenestemenn- og kvinner» (§ 1-2). I samme forskrift fremgår at testing og vurdering av studentenes kompetanser skal skje i tråd med målsetningene i UH-loven (§ 5-1). PHS' kvalitetssikringssystem fastslår at hovedformålet med kvalitetsarbeidet ved skolen er at studentene skal oppnå læringsutbyttene i studiet. Særlig skal utdanningens profesjonsforankring, relevans og studentenes læringsutbytte, vektlegges (Politihøgskolen, 2016a). Læringsutbyttene for politiutdanningen som helhet, og de overordnede målsetningene for hvert hovedområde, herunder etterforskning, fremgår i rammeplanen, mens de spesifikke læringsutbyttene for hvert emne fremgår av fagplanen (Politihøgskolen, 2014; 2015, s. 69-

71)³. Både rammeplan og fagplan fastslår at vurderingsformene i studiet skal danne grunnlag for å kunne vurdere studentenes kompetanse ut fra læringsutbyttene (Politihøgskolen, 2014, s. 8; 2015, s. 11). Samme målsetning formidles i ny rammeplan (Politihøgskolen, 2019b, s. 7).

2.4 Definisjon og begrepsforklaringer

Vurderingsfeltet er komplekst og omfattende, og inneholder mange begreper og underbegreper som delvis overlapper hverandre. For mitt formål og kontekst er mange av disse sentrale, og jeg har derfor sett det nødvendig å skrive et relativt omfattende delkapittel hvor jeg operasjonaliserer begrepene. Noen begreper vil bli omhandlet mer utførlig i teorikapittelet, og blir derfor kun kortfattet nevnt her.

Vurdering

Vurdering som fenomen kan forstås på ulike måter og benyttes i ulike sammenhenger. Helt grunnleggende handler vurdering om «menneskelig persepsjon, valg og sortering» (Vinge, 2012, s. 201). I didaktisk sammenheng knyttes begrepet primært til studentenes læring, enten som verktøy for å bidra til økt læring underveis i en læringsprosess, eller for å bedømme oppnådd læring gjennom avsluttende prøver og eksamener (Vinge, 2012, s. 201). Tidligere ble *vurdering* og *evaluering* brukt om hverandre som synonymer for eksamen, men i dag brukes *vurdering* om «all form for prøving av studenter», mens *evaluering* brukes om «ulike former for systemevaluering» (Raaheim, 2016, s. 19-20). Vinge skiller mellom ulike typer vurderingshandlinger. Vurderingen kan formidles verbalspråklig i form av skriftlige eller muntlige kommentarer, gjennom symbolske representasjoner som karakterer, eller gjennom kroppsspråk, for eksempel et nikk eller sukk (Vinge, 2012, s. 202). Jeg vil ikke gå nærmere inn på dette her, men nøyer meg med å påpeke at alle disse formene vil kunne gjøre seg gjeldende ved muntlig eksamen. Sadler definerer *vurdering* som «the process of forming a judgment about the quality and extent of student achievement of performance, and therefore by inference a judgment about the learning that has taken place» (Sadler, 2005, s. 177). I denne oppgaven forstås vurdering i tråd med ovennevnte. Vurdering er prosessen med å bedømme kvaliteten på studentprestasjoner ved å fatte slutninger over den læringen som har funnet sted.

³ Her henvises til rammeplan som var gjeldende på tidspunkt da jeg gjennomførte min datainnsamling

Summativ vurdering og eksamen

Summativ vurdering omtales ofte som vurdering av læring, og kan defineres som «vurdering ved endt utdanning som skal informere om elevens samlede kompetanse i et fag»

(Kunnskapsdepartementet, 2014, s. 39). Den finner sted etter at læringsaktivitetene i emnet er avsluttet og er en test av hva den lærende kan eller har lært (Raaheim, 2016, s. 21-22).

Eksamen i etterforskningsemnet er en summativ vurdering. Begrepet *eksamen* forstås i tråd med definisjonen som er lagt til grunn i forslag til ny UH-lov, og som også ble lagt til grunn ved utredning av ny lov i 1993: Med eksamen menes «... alle enkeltprestasjoner i form av prøver, oppgaver og avhandlinger m.m. som er tellende og føres på vitnemålet, eller som inngår som en del av slike eksamener ...» (NOU 2020:3, s. 199).

Læringsutbytte og kvalifikasjon

Læringsutbytte er beskrevet i kategoriene kunnskap, ferdigheter og generell kompetanse, og er «det en person vet, kan og er i stand til å gjøre som et resultat av en læringsprosess»

(Kunnskapsdepartementet, 2014, s. 37). En *kvalifikasjon* blir ofte betegnet som *formell kompetanse* og er «et formelt læringsutbytte på et visst nivå, godkjent av en instans og som kan dokumenteres» (Kunnskapsdepartementet, 2014, s. 35). I lys av denne forståelsen kan en kvalifikasjon nærmest sidestilles med en autorisasjon, ved at den er ment å «sjekke ut» de beskrevne, formelle læringsutbyttene. Kvalifikasjonsbegrepet brukes også i en bredere forståelse av flere av mine informanter, ved at de mer helhetlig fokuserer på hvilke kvalifikasjoner som trengs for å utføre etterforskningen i det virkelige liv. Jeg vil komme tilbake til det.

Kriteriebasert og normbasert vurdering – to ulike prinsipper for vurdering

I *normbasert vurdering* er målet å rangere studentene innbyrdes i et kull eller gruppe.

Studentprestasjonene sammenlignes med hverandre, heller enn med absolutte kriterier, slik at karakterene spres langs karakterskalaen etter en ønsket fordeling (Raaheim, 2016, s. 43-44).

ECTS er et normbasert system, hvor for eksempel de 10 % sterkeste kandidatene skal gis karakter A (Universitets- og høyskolerådet, 2008, s. 3). Innenfor et slikt relativt system vil det være vanskelig å sammenligne kandidater på tvers av tid og sted, fordi vurderingskriteriene er forankret i nivået på prestasjonene innad i en gruppe (Gynnild, 2013, s. 28).

Kriteriebasert vurdering har en kortere historie enn den normbaserte, og fikk særlig aktualitet etter innføring av kvalifikasjonsrammeverket, med sitt fokus på resultater av læringsprosessen

(Raaheim, 2016, s. 44). Kriteriebasert vurdering innebærer at «sensor skal vurdere en prestasjon basert på fastsatte kriterier og ikke det relative resultatet innenfor en gruppe eller et kull» (Meld. St. 16 (2016-2017), s. 56). Intensjonen er at vurderingen skal gjenspeile studentenes kompetanse (Gynnild, 2013, s. 39). I hvilken grad kompetansen er oppnådd, fastslås gjennom karaktersystemet. Dersom en prestasjon tilfredsstillende kriteriene for en karakter, skal karakteren gis uavhengig av hvordan fordelingen av de øvrige karakterene i eksamenskullet er (Universitets- og høyskolerådet, 2008, s. 3). Samtidig er det en forventning fra departementet om at karakterfordelingen, over tid og for et stort antall kandidater, skal være i samsvar med ETCS-skalaen (Universitets- og høyskolerådet, 2008, s. 3). Dermed er det norske karaktersystemet både absolutt, i forstand av å være kriteriebasert, og relativt, gjennom en forventning om normfordeling over tid. Bruken av karakterskalaen ligger dermed i spenningsfeltet mellom de to ulike prinsippene for vurdering (Gynnild, 2010, s. 27-28). I senere kapitler om funn og drøfting vil jeg komme tilbake til hvordan vurdering foregår i lys av disse perspektivene.

Kriterier og standarder

Kriterier er spesifikke kvaliteter (egenskaper) ved et læringsutbytte eller en kompetanse, og angir hva som skal inkluderes og vektlegges i vurderingen (Gynnild, 2010, s. 7). Sadler definerer kriterier som «attributes or rules that are useful as levers for making judgments» (Sadler, 2005, s. 179). Kriterier kan forstås både som noe eksplisitt, uttalt, nedskrevet og tilgjengelig, og som noe implisitt, personlig, uttalt og som vanskelig lar seg verbalisere (Sadler, 2009a, s. 819; Vinge, 2012, s. 202). Særlig i tilfeller hvor sensor skal foreta en helhetlig vurdering av komplekse forhold, vil vedkommende typisk benytte seg av (et spekter av) implisitte kriterier ut fra sitt profesjonelle skjønn og egen erfaring (Gynnild, 2010, s. 10). I etterforskningsemnet er det utarbeidet eksplisitte faglige vurderingskriterier. *Standarder* viser til nivået på prestasjonen (Gynnild, 2010, s. 7). De er sosialt forankret og stabile, i alle fall over en viss tidsperiode. De er eksplisitte, og «...the property of the academy as a collective, and are not determined or held privately by individual teachers and course teams» (Sadler, 2009a, s. 819). Standardene i etterforskningsemnet består av de kvalitative karakterbeskrivelsene. Begrepene vil bli omhandlet ytterligere i teorikapittelet.

Kompetanse og vurdering av kompetanse

Kompetansebegrepet brukes i mange ulike sammenhenger og blir definert og forstått ulikt av ulike aktører, avhengig av kontekst (NOU 2018:2, s. 14). Hvordan vi forstår kompetanse får

betydning for hvilke vurderingsformer vi velger og hva vi mener skal verdsettes og vektlegges i vurderingene (Lizzio & Wilson, 2004, s. 469). Særlig to aspekter synes å gå igjen på tvers av definisjoner av kompetanse, nemlig at de integrerer kunnskap, ferdigheter og holdninger, og at de refererer til en bestemt jobbsituasjon- eller kontekst (se for eksempel EC, 2007, s. 3; Lizzio & Wilson, 2004, s. 470; Meld. St. 16 (2015-2016), s. 3; NOU 2020:2, s. 29). Vurdering i kompetansebasert utdanning handler om å fatte slutninger om en students kompetanse (Baartman et al., 2007, s. 116, 118). Harlen skriver at «assessment in the context of education is a process of deciding, collecting and reasoning from evidence about learners' knowledge and skills» (Harlen, 2005, s. 263). Mulder skiller mellom oppfatninger av kompetanse som enten *competence based* eller *competence oriented*. Førstnevnte er orientert mot læringsutbyttene, og hvorvidt en eller flere av disse er oppfylt, mens sistnevnte fokuserer på (de presterte) kompetansenes relevans for arbeidslivet for aktuell profesjon. Innen kompetanseorientert vurdering vektlegges og verdsettes de kompetanser som har tilstrekkelig prediktiv validitet for suksess i arbeidslivet, altså kompetanser egnet til å gi informasjon om senere jobbprestasjoner (Mulder, 2014, s. 6). Som jeg vil komme tilbake til, synes begge oppfatninger å gjøre seg gjeldende ved vurdering av etterforskningskompetanse. I denne oppgaven forstås kompetanse innenfor rammen av kompetansebasert vurdering, ved at det er studentenes kompetanser som skal vurderes. Jeg ønsker ikke å avgrense kompetansebegrepet ytterligere, fordi det nettopp er informantenes oppfatninger av kompetanse som vil sette rammer for hva som vektlegges og verdsettes i vurderingene.

Analytisk og holistisk vurdering

I kompetansebasert vurdering kan man benytte ulike tilnærminger til vurdering i spennet mellom analytisk og helhetlig (holistisk) vurdering – som kan ses som to ytterpunkter av en skala. Fautley omtaler disse som henholdsvis *bottom-up* og *top-down*-strategier (Fautley, 2010). Som jeg vil komme tilbake til i drøftingskapittelet, er det min oppfatning at elementer fra begge tilnærminger anvendes i etterforskningsemnet. *Analytisk vurdering* innebærer vurdering av enkeltdeler ved prestasjonen ved bruk av eksplisitte kriterier (Sadler, 2005, s. 179). Sensor vil typisk kunne huke av, eller gradere, hvorvidt studenten har oppnådd hver av de delene som ønskes målt. Metoden er ment å sikre en rettferdig og pålitelig vurdering av studentenes prestasjoner, men kritiseres for å være uegnet ved vurdering av komplekse forhold som vanskelig lar seg bryte ned i biter av kunnskap og ferdigheter (Vinge, 2012, s. 204). Den motsatte tilnærmingen omhandles gjerne som helhetlig eller holistisk (Sadler, 1989). Ved *helhetlig vurdering* forstås kvalitet som et integrert konsept, hvor mange ulike

kriterier kan ligge til grunn for helhetsbedømmingen av prestasjonen. Her foreligger ingen forhåndsdefinerte beskrivelser av hver kompetanse, eller læringsutbytte, som forløpende skal vurderes. Kriteriene appliseres i etterkant av prestasjonen, ofte i form av å være implisitte og basert på sensors profesjonelle skjønn. De blir synlige først når de brukes. Innenfor denne tilnærmingen vil prestasjoner med forskjellige styrker og svakheter kunne bedømmes likt på bakgrunn av vidt forskjellige kriterier (Vinge, 2012, s. 204-205). For eksempel vil en kandidat som viser forståelse for etterforskningsmetodikk gjennom noen få men velbegrunnede eksempler, kunne oppnå like godt resultat som en student som inkluderer et bredere spekter av etterforskningsskritt.

Autentisk vurdering

Autentisk vurdering handler om å tilstrebe likhet i vurderingssituasjonen sammenlignet med tilsvarende situasjon i det virkelige liv. Vurderingsformene bør speile situasjoner og problemstillinger som en ferdigutdannet kandidat kan forvente å møte i arbeidslivet (Raaheim, 2016, s. 36). Autentisitet er et sentralt begrep i politiutdanningen og ved vurdering av etterforskningskompetanse, og vil bli ytterligere omhandlet i blant annet teorikapittelet.

Validitet og reliabilitet i vurdering og eksamen

Validitet og reliabilitet er viktige begreper for kvalitet i vurdering og forskning. I oppgaven opererer jeg med begrepene på begge nivåer, men i ulike deler. I denne delen omhandles de i vurderingssammenheng. Jeg vil redegjøre mer utførlig for begrepene i teorikapittelet, og gir derfor kun en kortfattet forklaring på dem her. *Validitet* er et uttrykk for gyldighet, og handler om en tests evne til å måle det den er ment å måle (Raaheim, 2016, s. 64). Reliabilitet handler om stabilitet i vurderingsresultatet, altså i hvilken grad samme produkt vurderes likt på tvers av sensorer (Raaheim, 2016, s. 63). Reliabilitet forstått som grad av enighet mellom sensorer, omtales ofte som *interbedømmer reliabilitet* eller *bedømmer-reliabilitet*. Jo likere vurdering, jo høyere interbedømmer reliabilitet (Raaheim, 2000, s. 204). Andre begreper brukt om det samme er *sensorsamsvar*, *sensorreliabilitet* og *vurderer-reliabilitet* (se for eksempel Asmyhr, 2011; Bjølseth, Havnes & Lauvås, 2011; Rye, 2014). I min sammenheng handler interbedømmer reliabilitet om hvorvidt sensorerne er enige med hverandre – ikke bare i karakterbedømmingene, men også i det som ligger til grunn for disse.

2.5 Forskning på feltet

Utdanningsdirektoratets eksamensgruppe har sammenstilt kunnskapsgrunnlaget vi har om eksamen i dag, og skriver at det generelt finnes lite forskning om eksamen – både nasjonalt og internasjonalt. De studiene som finnes har i hovedsak blitt gjennomført i et målingsperspektiv med vekt på blant annet sensorreliabilitet og konstruktvaliditet (Eksamensgruppa, 2019, s. 8). Det er forenlig med det jeg selv har funnet i arbeidet med å få oversikt over forskning på feltet. Arbeidet har funnet sted gjennom hele forskningsprosessen, og består i litteratursøk i databaser, søk på internett, gjennomgang av litteraturlister og lesning. Jeg har søkt på en rekke søkeord, både norske, svenske, danske og engelske, og i ulike databaser, herunder Oria, Idunn, Google Scholar, NORA, Academic Search Premier, Eric, Libris og Rex. Jeg har også lagret mange søk og søketreff, og lagt inn varsler i databasene for eventuelle nye publiseringer tilknyttet bestemte søkeord. Eksempler på søkeord har vært *vurdering, eksamen, muntlig eksamen, sensor, summativ vurdering, vurderer-reliabilitet, interbedømmer reliabilitet, sensorreliabilitet, sensorsamsvar, karaktersetting, holistisk vurdering, summative assessment, oral exams, oral examinations, inter-rater reliability, student evaluation, assessment consistency, higher education exam, muntlig exam, utvärderingsmetodik, utvärdering og muntlig eksamen*. Jeg har også fått hjelp fra bibliotekar til dette arbeidet.

Jeg har funnet en rekke studier hvor man har undersøkt grad av sensorsamsvar i karakterbedømming, men lite om hva som danner grunnlaget for vurderingene og hva som vektlegges i samspillet mellom sensorer når karakter skal settes. Jeg finner i det hele tatt lite forskning på vurdering av muntlig eksamen. Det er i tråd med Dobson, som skriver at det er gjort lite forskning både i Norge og internasjonalt for å klargjøre hva som foregår under muntlig eksamen (2007, s. 137). Internasjonalt har jeg særlig funnet forskning fra Nederland, men også fra andre land, som Danmark, Storbritannia, Australia og USA – i hovedsak rettet mot sensorsamsvar ved karaktersetting. Forskningen er i stor grad fra samfunnsvitenskapelige fag, helsefag og naturfag. Det er ikke forsket på temaet ved PHS, og jeg har ikke funnet studier om temaet fra politiutdanninger i andre land. Både Harlen (2005) og Bloxham (2009) har gjennomgått britiske forskningsresultater knyttet til sensorsamsvar, og konkluderer med at det ikke er sterk basis i forskning til å hevde at eksamensbesvarelser bedømmes likt. Det er i tråd med mine undersøkelser omkring forskning på feltet, som i all hovedsak finner lavt sensorsamsvar (se for eksempel Asmyhr, 2011; Bjølseth et al., 2011; Gynnild, 2010; Harlen, 2005; Raaheim, 2000; Rye, 2014).

I dette kapittelet har jeg redegjort for prosjektets kontekst, historikken som har ført frem til dagens karaktersystem, sentrale styringsdokumenter for vurdering som gjelder for PHS, sentrale begreper og forskning på feltet. I det følgende vil jeg beskrive fire teoretiske perspektiver.

3. Teoretiske perspektiver

Først vil jeg omhandle et femdimensjonalt rammeverk for autentisk vurdering utviklet av nederlandske forskere. Deretter vil jeg redegjøre for validitet og reliabilitet, som er sentrale kvalitetsindikatorer i all vurderingssammenheng. Videre vil jeg omhandle noen retoriske perspektiver med fokus på argumentasjon og overtalelsesmidler, som jeg mener er relevante å se i lys av mine funn. Som jeg vil komme tilbake til, synes slike forhold å være av betydning ved karakterbedømmingen i kommisjonene. Til slutt vil jeg redegjøre for skjønnsutøvelse i vurdering, fordi skjønn og subjektivitet alltid vil være gjeldende i kompetansebasert og autentisk vurdering – herunder ved vurdering av etterforskningskompetanse. I drøftingskapittelet vil jeg drøfte de teoretiske perspektivene i lys av problemstilling og funn.

3.1 Autentisk vurdering

De tre nederlandske forskerne Gulikers, Bastiaens og Kirschner har utviklet et rammeverk for autentisk vurdering som består av fem dimensjoner, som hver kan variere i grad av autenticitet (Gulikers, Bastiaens & Kirschner, 2004). Rammeverket er basert på omfattende litteraturstudier av autentisk vurdering, og er ment å være et verktøy for design av autentiske vurderingsopplegg med profesjonell praksis som utgangspunkt. Det knytter modellen tett opp til oss som jobber med profesjonsutøvelse og som skal teste profesjonsutøvelse gjennom muntlig eksamen i etterforskning. Gjennom Studietilsynsforskriften (§ 2-2) er PHS underlagt lovpålagt krav om autenticitet i utdanningen. Målet med rammeverket er å kaste lys over begrepet (vurderings)autenticitet og å gi retningslinjer for hvordan en kan implementere autentiske elementer i kompetansebasert vurdering (Gulikers et al., 2004, s. 70). Før jeg beskriver hver dimensjon, vil jeg omhandle begrepet autenticitet.

Autenticitet har blitt et viktig begrep innen nyere former for vurdering. Det hersker ulike oppfatninger av hva autentisk vurdering er, og hvilke kompetanser som skal testes i en slik

vurdering. Forståelsen spriker i spennet fra å omfatte testing av virkelighetsnære kognitive kompetanser (tankeprosesser), til å kunne anvende kunnskaper, ferdigheter og holdninger på virkelighetsnære problemstillinger, til å (også) inkludere testing av metakognitive og sosiale kompetanser som refleksjon, kommunikasjon og samarbeid (Gulikers et al., 2004, s. 67-71). Gulikers et al. bruker følgende definisjon av autentisk vurdering:

An assessment requiring students to use the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in professional life. The level of authenticity of an assessment is thus defined by its degree of resemblance to the criterion situation (Gulikers et al., 2004, s. 69).

Av definisjonen følger at grad av autenticitet henger sammen med grad av likhet mellom vurderingssituasjonen og tilsvarende oppgaveutførelse i arbeidslivet. Imidlertid er det en utfordring at autenticitet er subjektivt og preget av oppfatninger. Det er ikke sikkert at oppfatningene er like på tvers av fagansatte og studenter – eller innad i lærer- og studentgrupper. Ulike forståelser kan gi uheldige utslag for læring og i vurdering, fordi studentene kan ende opp med å rette læringsinnsatsen mot noe annet enn det de blir vurdert i. Studentene er ofte strategisk i sitt læringsarbeid og retter sin tidsbruk mot hva de tror de skal vurderes i og hva de tror vil gi dem gode karakterer. For eksempel vil deres forståelser og oppfatninger av vurderingskriteriene kunne ha betydning for hva de lærer og hva de presterer til eksamen (Gibbs, 2006, s. 22; Gulikers et al., 2004, s. 68). Autentisk vurdering (og opplæring) har dermed en sentral plass for å fremme autentisk læring (Gulikers et al., 2004, s. 67). I det følgende vil jeg redegjøre for dimensjonene for autentisk vurdering.

3.1.1 Fem dimensjoner for autentisk vurdering

Følgende fem dimensjoner har betydning for autenticiteten i vurderingen: oppgaven, den fysiske konteksten, den sosiale konteksten, vurderingsresultatene og vurderingskriteriene (Gulikers et al., 2004, s. 71-75). Hver dimensjon kan variere i grad av autenticitet, og er ment å utvide og konkretisere forståelsen av hva autenticitet er. Jeg anser det som relevant å se alle dimensjonene i lys av problemstillingen, de tre første mest som et bakteppe og de to siste mer direkte opp mot vurderingene. Jeg vil komme tilbake til det.

Oppgaven (what do you have to do): En autentisk oppgave krever at studentene integrerer de kunnskaper, ferdigheter og holdninger som kreves ved tilsvarende oppgaveutførelse i

profesjonell praksis. For meg er det aktuelt å se på hva oppgaven til etterforskningseksamen ber studentene om å besvare, og sammenholde det med hvordan prestasjonene vurderes og hvordan tilsvarende oppgaveutførelse skjer i det virkelige liv. I en autentisk vurdering bør oppgaven speile den kompleksitet som gjelder for samme type oppgaveutførelse i arbeidslivet. I tillegg bør den oppfattes som relevant og meningsfull ved at studentene ser dens overføringsverdi til det virkelige liv. En utfordring med sistnevnte er at studentenes oppfatninger varierer – både på tvers av studenter og på tvers av tid. For eksempel vil oppfatningene kunne endres etter hvert som studentene blir mer erfarne. I mitt prosjekt har jeg ikke samlet data om studentenes oppfatninger. Nore og Lahn skriver at denne dimensjonen blant annet omfatter strukturelle krav til prøven, som tidsspenn (timer, dager eller uker) og bruk av hjelpemidler (Nore & Lahn, 2015, s. 148).

Fysisk kontekst (where do you have to do it): I en autentisk vurdering bør konteksten være gjenkjennbar sammenlignet med tilsvarende oppgaveutførelse i profesjonell praksis. En datasimulasjon kan være et godt eksempel på en høy-gjenkjennbar situasjon. For en autentisk fysisk kontekst bør antall og typer tilgjengelige ressurser være i tråd med ressurstilgangen i det virkelige liv. I profesjonell praksis vil slike ressurser inneholde både relevant og irrelevant informasjon for oppgaveløsningen, noe som også bør gjenspeiles i vurderingen. Også tidsaspektet bør være realistisk sammenlignet med arbeidslivet. Eksamener gjennomføres normalt over en begrenset periode, mens tilsvarende aktivitet i arbeidslivet gjerne er fordelt over en lenger periode. Ifølge Wiggins bør ikke en autentisk vurdering avhenge av urealistisk og vilkårlig tidsbegrensning (Wiggins, 1989).

Sosial kontekst (with whom do you have to do it): I autentisk vurdering er det viktig at de sosiale prosessene er gjenkjennbare sammenlignet med tilsvarende situasjon i arbeidslivet. Dersom situasjonen i det virkelige liv krever samarbeid, må vurderingen også involvere samarbeid. Dersom den reelle situasjonen normalt sett løses individuelt, bør vurderingen også være individuell.

Vurderingsresultat (what has come out of it): Vurderingsresultatet er produktet av oppgaven, i mitt tilfelle eksamensprestasjonen. Nore og Lahn kaller den det dokumenterte utbyttet (2015, s. 149). I rammeverket karakteriseres et autentisk vurderingsresultat av flere elementer. Blant annet bør det være en aktivitet som studentene kan bli bedt om å utføre i det virkelige liv, innrettet slik at det kan fattes gyldige slutninger om studentenes underliggende kompetanser.

Ofte vil det ikke være mulig å demonstrere alle relevante kompetanser i en enkelt test. For å kunne komme til gyldige og rettferdige slutninger om en students kompetanse, bør en autentisk vurdering derfor involvere en rekke oppgaver og indikatorer for læring (Darling-Hammond & Snyder, 2000, s. 527-528). Studenter bør gis anledning til å presentere sitt arbeid til andre, muntlig eller skriftlig, fordi de ved å forsvare sitt arbeid kan bevise sin kompetanse og mestring.

Vurderingskriterier og standarder (how does what you have done have to be evaluated or judged): Kriterier er de karakteristika eller egenskaper ved prestasjonen som verdsettes og tillegges vekt i vurderingen (vurderingskriteriene), mens standarder er forventet prestasjonsnivå for hvert karakterspenn (karakterbeskrivelsene) (Arter & Spandel, 1992). Det å utarbeide kriterier og gjøre dem eksplisitte og transparente for de lærende i forkant av vurderingssituasjonen er viktig i autentisk vurdering, fordi det styrer læringen. Også i det virkelige liv vil det ofte være slik at de ansatte vet hvilke kriterier deres prestasjon eller produkt blir vurdert etter. Dette impliserer at autentisk vurdering krever kriteriebasert vurdering. Noen kriterier bør springe ut av hvilke karakteristika eller krav som kreves av tilsvarende produkt eller prestasjon i det virkelige liv. Kriteriene bør også rettes mot utvikling av de relevante profesjonelle kompetanser som oppgaveløsningen krever. Gulikers et al. (2004) argumenterer også for at det er et gjensidig forhold mellom kriteriedimensjonen og de fire andre dimensjonene i rammeverket.

3.2 Validitet og reliabilitet i vurdering og eksamen

Validitet og reliabilitet er sentrale kvalitetskriterier innen kompetansebasert autentisk vurdering. *Validitet* er et uttrykk for gyldighet, og handler om en tests evne til å måle det den er ment å måle (Raaheim, 2016, s. 64). Å bedømme validitet i eksamen innebærer å ha et blikk på hele prosessen, fra oppgaveutvikling til vurdering, og til måten vurderingene blir anvendt på (Eksamensgruppa, 2019, s. 28). Det finnes mange ulike definisjoner av validitet, men det synes likevel å være bred enighet om at en tolkning er valid dersom den støttes av tilstrekkelige bevis (Baartman et al., 2007, s. 119; Kane, 2016, s. 198). Moss et al. skriver at validitetsbedømminger handler om hvorvidt en tolkning, avgjørelse eller handling oppfattes som fornuftig, og hvilke bevis, resonnementer eller kriterier som ligger til grunn for den (Moss, Girard & Haniford, 2006, s. 109). Validiteten har dermed sterk tilknytning til premissene som legges til grunn for bedømmingen, og deres logiske sammenheng med

karakterslutningen. Hvilke krav som stilles til gyldighet vil variere avhengig av hvilket fokus man har og hvilken vurderingsform som benyttes (Raaheim, 2016, s. 64). Det kan eksemplifiseres ved å gjøre et skille mellom *indre* og *ytre validitet*. Ved summativ vurdering vil den indre validiteten påvirkes av i hvilken grad eksamen faktisk tester det studentene er ment å ha øvd på, mens den ytre validiteten handler om i hvilken grad testresultatet har overføringsverdi til senere situasjoner (Raaheim, 2016, s. 64-65).

I det femdimensjonale rammeverket argumenterer forskerne for at den fremste årsaken til at man bør benytte autentiske vurderingsoppgaver, er deres betydning for konstrukt- og konsekvensvaliditet (Gielen, Dochy & Dierick, 2003; Gulikers et al., 2004, s. 68). *Konstruktvaliditet* handler om hvorvidt en vurdering måler det den er ment å måle, mens *konsekvensvaliditet* viser til vurderingenes ønskede og uønskede effekter på undervisning og læring. Læring og vurdering er to sider av samme sak, og har stor påvirkning på hverandre. For å styre studentenes læringsinnsats mot aktiviteter egnet til å utvikle ønskede og relevante kompetanser, må undervisning og vurdering være tilpasset hverandre – i tråd med Biggs' constructive alignment (Biggs, 1996; Gulikers et al., 2004, s. 68). Validiteten vil også reduseres dersom testsituasjonen fremmer (konsekvensen) ekstrem nervøsitet, slik at den i realiteten blir en stressmestringstest heller enn en test av læring (Eksamensgruppa, 2019, s. 2; Raaheim, 2016, s. 65). Konsekvensvaliditet handler også om de følger vurderingsresultatet får for videre muligheter i studier og arbeidsliv (Baartman et al., 2007, s. 123).

Særlig tre kriterier har betydning for validiteten i kompetansebasert vurdering. For det første må eksamensoppgaven reflektere den kompetansen, altså de læringsutbytter, som studentene skal vurderes i. For det andre må vurderingen knyttes til autentiske oppgaver som representerer virkelighetsnære og relevante problemstillinger for aktuell profesjon. For det tredje må tankeprosessene som oppgaveløsningen krever i det virkelige liv, være påkrevd for å løse eksamensoppgaven (Gielen et al., 2003). Basert på disse kriteriene har autentisk kompetansebasert vurdering høyere validitet i måling av kompetanser enn såkalte objektive eller tradisjonelle tester, hvor svaralternativene er forhåndsdefinerte og hvor formålet er å reproducere et bestemt materiale (Gulikers et al., 2004, s. 68). Autentisk vurdering kan også ha betydning for *prediktiv validitet*. Det er en form for gyldighet som handler om å kunne forutsi studentenes evne til å prestere i fremtidig profesjonell praksis (Gulikers et al., 2004, s. 75). Økt samsvar mellom vurderingen og tilsvarende oppgaveløsning i arbeidslivet, vil trolig føre til økt autentisk læring, og dermed til økt prediktiv validitet i vurderingen (Gulikers et al.,

2004, s. 75).

Messick har utviklet et rammeverk for validitet som har vært mye brukt innen tradisjonell testing (Messick, 1987, 1994, 1995). I rammeverket bruker han *konstruktvaliditet* (construct validity) som overordnet validitetsbegrep, definert som «an integration of any evidence that bears on the interpretation or meaning of the test scores» (Messick, 1987, s. 10). Av det overordnede begrepet skiller han ut seks aspekter som validitet kan bedømmes ut fra, herunder innholdsvaliditet, strukturvaliditet og konsekvensvaliditet. I et senere utviklet rammeverk for kompetansebasert vurdering (CAP), fant man at Messicks validitetsbegreper for tradisjonell testing i stor grad er forenlige med kvalitetskriterier for kompetansebasert vurdering (Baartman et al., 2007, s. 122). Blant annet relateres *innholdsvaliditeten* til kvalitetskriteriene *autentisitet* og *rettferdighet*. Disse validitetsbedømmingene handler om hvorvidt eksamensoppgaven og dens kontekst er virkelighetsnær, om studentene får anledning til å demonstrere hva de har lært og om kompetansene vurderes som en integrert del av den større helheten de realistisk vil inngå i (Baartman et al., 2007, s. 122-123). *Strukturell validitet* knyttes til kvalitetskriteriene *rettferdighet* og *transparens*. Denne formen for validitet fokuserer på hvorvidt kriteriene som vektlegges i vurderingen medfører urettferdigheter for studentene, og om hvorvidt bruken av kriterier er transparent (Baartman et al., 2007, s. 123-125).

Reliabilitet er et uttrykk for pålitelighet i testresultater. Det handler om «...the degree to which the same results would be obtained on a different occasion, in a different context, or by a different assessor» (Baartman et al., 2007, s. 118). I standardisert testing handler reliabilitet om stabilitet i testskåren, hvor en reliabilitetskoeffisient under 0,80 vanligvis ikke anses som akseptabel (Raaheim, 2016, s. 63). Imidlertid er det ikke slik at reliabilitet er betinget av objektivitet og standardisering, og at en avkrysningsprøve nødvendigvis er mer reliabel enn en muntlig eksamen. Ingen metode er iboende ureliabel og enhver metode kan være tilstrekkelig reliabel, forutsatt at testformen er egnet til å prøve det som skal måles. Det avgjørende er hvordan vurderingen innrettes og hvilke valg som gjøres, for eksempel hvor mye tid studentene gis til å bevise sin kompetanse og antall vurderinger kompetansen bedømmes ut fra (Van Der Vleuten & Schuwirth, 2005, s. 312). Kontekstuelle faktorer, som hvem som vurderer og hvilken vurderingsform som benyttes, utgjør viktige betingelser for reliabiliteten (Raaheim, 2016, s. 63).

Innen kompetansebasert vurdering vil testresultatet være preget av subjektive vurderinger, og spørsmål knyttet til reliabilitet vil dreie seg om i hvilken grad en studentprestasjon blir vurdert (tilnærmet) likt på tvers av sensorer. Reliabilitet forstått som grad av enighet mellom sensorer, omtales ofte som *interbedømmer reliabilitet*, *bedømmer-reliabilitet* eller *sensorreliabilitet*. Jo likere vurdering, jo høyere interbedømmer reliabilitet (Baartman et al., 2007, s. 118; Raaheim, 2000, s. 204-205). Schuwirth og van der Vlauten skriver at det avgjørende i en reliabilitetsanalyse er «...the reproducibility of the decisions that are made on the basis of the results of the tests» (Schuwirth & Van Der Vleuten, 2004, s. 806). Reproduserbarhet, eller generaliserbarhet, handler om hvorvidt vurderingssituasjonen gjør det mulig å trekke mer generelle konklusjoner om en students kompetanse (Baartman et al., 2007, s. 121). For mitt formål er et grunnleggende spørsmål knyttet til reliabilitet hvorvidt sensorene, og kommisjonene, er enige med hverandre i sine vurderinger av de to studentenes prestasjoner.

Validitet og reliabilitet henger sammen. I den grad vurderinger er upålitelige, vil også gyldigheten svekkes (Sadler, 2009b, s. 162).

3.3 Retorikk og bruk av overtalelsesmidler i vurdering

I det følgende vil jeg redegjøre for noen perspektiver hentet fra retorikken. Jeg ble mer og mer oppmerksom på denne innfallsvinkelen gjennom arbeidet med å bryte ned og analysere empirien, fordi jeg gjorde stadige oppdagelser av at forhold ved kommunikasjonen og argumentasjonen syntes å ha betydning for bedømmingen i kommisjonene. Jeg anså det som interessant og relevant å se funnene i lys av retorikken og dens ulike former for overtalelse. Perspektivet er særlig aktuelt i lys av forskningsspørsmål nummer 2 (hvordan vektlegges de ulike vurderingene i samspillet når karakter skal settes).

Retorikk er den eldste kommunikasjons- og språkvitenskapelige tradisjonen vi kjenner. Den strekker seg tilbake til antikken, hvor fremveksten av et delvis demokratisk styresett i Athen gjorde det viktig å kunne uttrykke seg godt og å påvirke tilhørere. I tillegg til Aristoteles (384-322 f.Kr.), var Cicero (106-43 f.Kr.) og Quintilian (35-100 e.Kr.) sentrale personer for retorikkens utvikling, og mange av deres teorier er fortsatt aktuelle. I dag er retorikk en viktig del av både språkopplæring og utdanningssystemet i mange land. Den brukes som et kritisk redskap, og fokuserer på rammer rundt, og betingelser for, kommunikasjon. Den har grensedragninger mot psykologi, sosiologi, statsvitenskap og filosofi, blant annet ved at den

ser på hvordan mennesker blir overtalt eller overbevist, på språkets rolle i sosiale og politiske prosesser og på språkets betydning for hvordan vi forstår verden rundt oss (Svendsen & Grue, 2019).

En retorisk ytring er forsøk på å overbevise, eller overtale, gjennom kommunikasjon (Johannessen, Rafoss & Rasmussen, 2018, s. 186). Det handler om å skape troverdighet (Eide, 2006, s. 8, 25). Quintilian definerer retorikk som «kunsten å tale godt», hvor godt sikter til talens virkningsfullhet, det vil si evnen til å overbevise (Eide, 1999, s. 31). Aristoteles definerer retorikk som: «evnen til i enhver sak å se hvilke muligheter vi har til å overtale» (Eide, 2006, s. 27). Retorikkens oppgave ligger innenfor områder hvor vi drøfter saker som synes å ha (minst) to sider. Dens oppgave er ikke overtalelsen i seg selv, men å se overtalelsesmulighetene i hvert tilfelle. Målet er å hjelpe tilhørerne med å fatte avgjørelser (Eide, 2006, s. 25, 30, 103). Eide påpeker at oversettelse av Aristoteles' definisjon fra gresk til norsk, forenkler meningsinnholdet og gjør det noe unyansert. Det greske ordet *dynamis*, som i definisjonen er oversatt til evne, betyr også makt, kompetanse, talent, potensialitet og iboende muligheter. Ordet *theorein*, som er oversatt til å se, betyr også å studere og å vurdere (Eide, 2006, s. 8). Ordet *peithein*, som i definisjonen er oversatt til å overtale, betyr også å overbevise. Det favner både det å overtale noen til å gjøre noe, og å overbevise noen til å tro eller innse noe (Eide, 1999, s. 27; Slaattelid, 2006, s. 389).

Enten målet for talen er å anbefale, fraråde, rose, kritisere, forsvare eller anklage, må taleren kjenne til de retoriske premissene for saken. Dersom målet er å anbefale noe, vil taleren typisk argumentere for det han mener er det beste valget og fraråde det han mener er det dårligste. For å kunne gjøre det, må han ha kjennskap til argumentene egnet til å overbevise publikum om hvorfor et valg er bedre enn et annet. Talerens kunnskap om relevante premisser har betydning ikke bare for talens innhold og argumentasjon, men også for publikums oppfatning av hans troverdighet og for hans evne til å vekke de ønskede følelser hos dem han snakker til (Eide, 2006, s. 35-52). Det henspiller på de tre deler som Aristoteles deler talen inn i: taleren, emnet for talen og den taleren henvender seg til. Det er sistnevnte, tilhøreren, som skal overbevises (Eide, 2006, s. 35).

3.3.1 Overtaleelsesmidlene (appellformene)

For å overbevise sitt publikum har taleren tre overtaleelsesmidler til disposisjon, også kalt bevismidler eller appellformer. Disse er *etos*, *patos* og *logos*. *Etos* er rettet mot talerens personlige karakter, *logos* mot argumentasjonen i talen og dens beviskraft, og *patos* mot sinnstilstanden tilhøreren settes i gjennom talen (Eide, 2006, s. 27-28). Overtalelsen skjer alltid gjennom (minst) en av disse tre, altså enten ved at publikum får et godt inntrykk av taleren, at saken blir sannsynliggjort eller bevist eller at tilhørerne blir påvirket på en bestemt måte (Eide, 2006, s. 203). Ifølge Cicero har taleren tre oppgaver, nemlig å belære, behage og bevege sitt publikum. Sett i lys av bevismidlene, kan det sies at taleren kan belære gjennom *logos*, behage gjennom *etos* og bevege gjennom *patos* (Andersen, 1995, s. 42). Forsøk på å overbevise skjer aldri fri for kontekst. Enhver retorisk analyse forutsetter at ytringer forstås i lys av konteksten de er uttalt i – både den smale (umiddelbare) og brede (kulturelle og historiske) (Johannessen et al., 2018, s. 188) Under følger en nærmere redegjørelse av hver appellform.

Etos viser til talerens karakter. Den har stor betydning for talens troverdighet og overbevisningskraft. Tilhørerne må ha tro på taleren før de tror på saken, og derfor må taleren ses som en viktig del av budskapet (Andersen, 1995, s. 35). Aristoteles skriver det så sterkt som at talerens karakter nærmest har avgjørende betydning for troverdigheten i talen (Eide, 2006, s. 27-28). Dette er også i tråd med Quintilian, som mener at det som angår taleren også angår saken (Quintilian referert i Andersen, 1995, s. 35). Det er inntrykket taleren klarer å gjøre på tilhørerne som utgjør hans *etos*, og ikke om han faktisk er troverdig (Eide, 2006, s. 9). For eksempel vil en taler som virker trygg og sikker fremstå mer troverdig enn en som skjelver og svetter, med mindre konteksten gjør at det er naturlig å skjelve og svette. *Etos* er i stor grad prissatt det inntrykket publikum har av taleren fra før, og derfor kan talerens omdømme, posisjon og anseelse ha betydning for evnen til å overbevise (Johannessen et al., 2018, s. 194-195). Talerens *etos* skapes gjennom *logos*, hvor tilhørernes oppfatninger av hans troverdighet følger av hans argumentasjon (Kjeldsen, 2002, s. 30). For eksempel kan man etablere god karakter gjennom å fortelle om egne prestasjoner, eller eget liv (Andersen, 1995, s. 35-36). Om tilhørerne får grunn til å mistenke at taleren fører dem bak lyset, vil ordene miste sin overbevisningskraft (Slaattelid, 2006, s. 390). Derfor må taleren fokusere ikke bare på at talens innhold skal være troverdig og overbevisende, men også på egen fremtoning (Eide, 2006, s. 103-104). Makt har stor påvirkning på en persons karakter. Makten en person besitter kan i seg selv gi vedkommende gjennomslagskraft, særlig innen de områder som

angår makten (Eide, 2006, s. 157).

Patos handler om de følelser ytringen fremkaller hos publikum. Følelser kan påvirke våre oppfatninger og beslutninger, og kan være avgjørende for hvorvidt en taler lykkes i å overtale sitt publikum (Johannessen et al., 2018, s. 197). Våre avgjørelser påvirkes nemlig av om vi er glad, sint, trist, lei eller om vi er vennlig eller fiendtlig innstilt, og vi bedømmer ting ulikt avhengig av hvilken sinnsstemning vi er i. Som et ledd i å overtale publikum, kan taleren derfor innrette talen slik at han vekker ønskede følelser i dem – positive eller negative. Det å sette noen i en bestemt sinnsstemning kan bidra til å skape troverdighet (Eide, 2006, s. 8, 9, 28, 104). Hvilke følelser man bør spille på avhenger av den konkrete konteksten og hva som er formålet med talen. Det avhenger også av karaktertrekk ved publikum, som deres alder og livssituasjon (Eide, 2006, s. 147-155).

Logos viser til selve argumentasjonen i talen og dens reelle eller tilsynelatende beviskraft. Publikum tror på argumenter som overbeviser, enten ved at de er sanne og motsetningsfrie, eller ved at de sannsynliggjør det saken innbyr til (Eide, 2006, s. 27-28; Slaattelid, 2006, s. 390). Taleren mobiliserer logos når han gjennom overbevisende argumenter viser hva som er sant eller sannsynlig (Andersen, 1995, s. 34). Mennesker er utstyrt med evne til å identifisere sannheten og det som ligner på sannhet, og de fleste godtar allment aksepterte standpunkter som tilstrekkelige til at noe anses sannsynliggjort eller bevist. Det innebærer at allment godtatte meninger kan brukes som argumenter for å påvirke tilhørerne til å konkludere i en ønsket retning (Eide, 2006, s. 195). I mange tilfeller vil deler av argumentene forbli usagt og følge implisitt, ofte fordi taleren legger til grunn at alle er kjent med dem og er enige om dem (Johannessen et al., 2018, s. 201).

3.4 Subjektivitet og skjønnsutøvelse i vurdering

Eisner tar utgangspunkt i ontologisk og prosessuell objektivitet når han analyserer det objektivitetsidealet som i lang tid har preget det amerikanske utdannings- og forskningsmiljøet. *Ontologisk objektivitet* innebærer å se ting slik de er, altså å oppleve dem i sin sanne, eller objektive, tilstand. Det man ser er det som finnes, og det kan ikke endres av subjektive oppfatninger. *Prosessuell objektivitet* innebærer utvikling av metoder, for eksempel selvrettende tester, som (søker å) eliminere personlig skjønn i vurderingen og beskrivelsen av en tilstand. Testingen er ment å resultere i identiske og sammenlignbare resultater gjennom

prosedyrer som definerer hvordan bestemte kompetanser skal måles (Eisner, 2017, s. 43-44, 50). Eisner peker på en rekke utfordringer ved testing som søker å oppnå prosessuell objektivitet. Blant annet vil det ikke være slik at konsensus nødvendigvis gir et riktig bilde av verden, altså kompetansen. Det viser bare at vi kan være enige. Ofte forenkles kvalitetene som måles (læringsmålene) gjennom slik testing, slik at bedømmingen fordrer lite eller ingen tolkning. Andre ganger gjøres klare avgrensninger i hva som skal vurderes, slik at sensorene står overfor få og lite komplekse valg som gjør det «lett» å vurdere og som typisk gir høyt sensorsamsvar (Eisner, 2017, s. 47).

Selv om objektivitet er et vesentlig tema innen vurderinger, er det ikke mulig for mennesker å være objektive. Våre oppfatninger påvirkes av våre erfaringer og kompetanser, synspunkt, fokus, språk og kontekst. Vårt øye er ikke bare en del av hjernen, men også av den tradisjonen vil tilhører (Eisner, 2017, s. 46). All vår kunnskap om verden fås gjennom erfaringer. Erfaringene bygger på tidligere erfaringer, og er formet av kultur, språk, behov og av alle de idéer, praksiser og hendelser som gjør oss menneskelige. Kunnskapen fra våre erfaringer struktureres i hjernen gjennom mentale skjema (Eisner, 2017, s. 46-47). Slike skjema brukes i læringspsykologien, blant annet hos Piaget, som metafor for å forstå måten hjernen organiserer kunnskapen på, og kan sies å definere rammene for vår forståelse av verden (Illeris, 2012, s. 22-23). Kunnskapsstrukturene skapes gjennom sosialisering, både på den profesjonelle arena og ellers. Hvordan man for eksempel oppfatter en undervisningsøkt, avhenger av mange faktorer. Blant annet vil den enkeltes interesser og oppfatninger av hva som er viktig, styre hva en ser etter og dermed hva en får med seg. Forhold som lærerens språklige formidlingsevne, samhandling med studentene, engasjement eller personlige stil, kan ha betydning. Hva vi ser avhenger av hva vi søker, og hva vi søker avhenger av hva vi kan *si* noe om. Et tomt sinn ser ingenting (Eisner, 2017, s. 46). Menneskers syn på verden har stor betydning for hva de er i stand til å si om den, og hva de opplever eller tar med seg fra en situasjon (Eisner, 2017, s. 7-8).

Standardiserte tester og definisjoner av hvordan man skal måle kompetanser, står i motstrid til det faktum at vi mennesker ikke kan være objektive. Slike tester representerer et kunnskapssyn om at det eksisterer en ytre virkelighet uavhengig av mennesket (Eisner, 2017, s. 43). Til dette spør Eisner hvordan vi noensinne kan vite at våre oppfatninger av virkeligheten korresponderer med verden slik den er. For å vite det må vi vite hvordan verden faktisk er, og hvis vi vet det trenger vi ikke å ha et syn på den. Og motsatt, siden vi ikke vet

hvordan verden faktisk er så kan vi ikke vite sikkert at vårt syn korresponderer med den. Eisner viser blant annet til Popper, som hevder at vi aldri kan verifisere sannheten i en påstand, kun gjendrive den, men aldri fullstendig (Eisner, 2017, s. 45).

Hvis vi som mennesker ikke kan være helt objektive, og i tillegg har ulike subjektive oppfatninger av verden: hvordan kan vi da vurdere om et produkt fremstår som troverdig eller ikke, og om vi skal si oss enig eller uenig med konklusjonen? Eisner peker på flere faktorer som være til støtte ved vurderingen, deriblant *sammenheng* (coherence) i argumentasjonen og mellom argumenter og slutning, og *konsensus* (Eisner, 2017, s. 53-58). Overført til mitt prosjekt kan disse faktorene ses i lys av sensorenes begrunnelser, både de skriftlige og muntlige. Er begrunnelsene sammenhengende, og hvordan skapes konsensus når karakter skal settes? *Sammenheng* viser til styrken i argumentene som presenteres. Gir argumentasjonen mening? Er det sammenheng mellom argumentasjon og konklusjon? Er det motstrid i argumentasjonen? Finnes det andre troverdige og plausible tolkninger? I vurderingen av sammenheng kan man benytte det Eisner kaller *structural corroboration* (strukturell underbygging). Det er en prosess som søker å validere eller støtte konklusjoner om et fenomen ved å demonstrere hvordan flere faktum eller betingelser underbygger konklusjonen (Eisner, 2005, s. 46). Prosessen krever at dataene er brutt ned og at de relevante informasjonsbiter egnet til å underbygge konklusjonen, er sammenstilt. Sammenstillingen kan vise samstemthet (eller motstrid) i dataene, eller gjentakende forekomster av lignende uttalelser eller hendelser egnet til å underbygge den (Eisner, 2017, s. 55). Prosessen kan sammenlignes med å legge et puslespill, hvor bildet gradvis viser seg gjennom biter som passer sammen. *Sammenheng* (coherence) kan også være villedende – bevisst eller ubevisst. Vi kan bli villedet av dem som har et ønske om å overbevise. Eller vi kan bli villedet av dem som er så overbevist om hva konklusjonen vil bli, at de kun fokuserer på informasjon egnet til å bekrefte egne tanker og oppfatninger, og overser informasjon som peker i andre retninger (Eisner, 2017, s. 54).

Konsensus kommer fra det latinske verbet *consentive*, som betyr å være enig. Det er en tilstand hvor funnene eller tolkningene som ligger til grunn, samsvarer med leserens (vurdererens) egne erfaringer eller de presenterte bevis. Konsensus handler om enighet, som igjen handler om overtalelse. Den strukturelle underbyggingen som er beskrevet over, kan være et viktig verktøy for å oppnå konsensus gjennom overtalelse. Konsensus er en form for det Eisner kaller gjentakende bekreftelse (multiplicative corroboration). Et eksempel er når

tolv jury-medlemmer er samstemte om, eller overbevist om, at bevisene er konsistente og sammenhengende nok til at de fjerner all rimelig tvil (Eisner, 2005, s. 46). Konsensus avler tillit til saklighet i dommen, og underbygger og bekrefter konklusjonene (Eisner, 2017, s. 45, 56). En utfordring er at vi mennesker ikke lar oss overbevise om det samme. Det noen finner overbevisende gjelder ikke for andre. Mennesker som har sterk motstand mot et tema lar seg vanskeligere overbevise enn mennesker som ikke har slik motstand. Og de som ikke har motstand er ofte mer åpen for å vurdere bevisene, eller argumentene. I tilfeller hvor leseren, eller vurdereren, må gjøre egne tolkninger for å konkludere, kan oppnåelse av konsensus være en krevende øvelse. Grad av konsensus henger tett sammen med kompleksiteten og subtiliteten i produktet. Det må forventes at grad av presisjon i oppnådd konsensus vil reduseres med økt kompleksitet. Konsensus innebærer ikke at man nødvendigvis har kommet frem til «sannheten», kun at man har oppnådd enighet. Hvilken grad av konsensus som er tilfredsstillende vil variere fra sak til sak, og blant annet avhenge av hva som bedømmes og omfanget av konsekvensene som beslutningen fører til (Eisner, 2017, s. 57-58).

Jeg har nå redegjort for fire teoretiske innfallsvinkler som jeg mener er relevante for problemstillingen. I det neste kapittelet vil jeg redegjøre for metodiske tilnærminger, valg og refleksjoner.

4. Metode

I det følgende vil jeg redegjøre for sentrale aspekter ved min forforståelse som jeg anser av betydning for undersøkelsen. Deretter beskriver jeg sentrale kvalitetskriterier innen kvalitativ forskning, før jeg redegjør for hvordan jeg har gått frem for å belyse problemstillingen – gjennom metodevalg, utvalg av informanter, gjennomføring av datainnsamling og transkribering. Til slutt oppsummerer jeg etiske betraktninger ved arbeidet.

I prosjektet tar jeg utgangspunkt i en induktiv tankemåte. I en induktiv tilnærming ønsker man å studere et fenomen mest mulig forutsetningsløst, gjennom en problemstilling som ikke tar utgangspunkt i bestemte teorier eller hypoteser (Halvorsen, 2008, s. 24, 73). Imidlertid er det ikke mulig å studere noe helt induktivt, eller forutsetningsløst, fordi forskerens subjektive forståelse av verden alltid vil være til stede og prege arbeidet (Postholm, 2010, s. 99). Derfor har jeg brukt mye tid til å reflektere over egen forforståelse og dennes betydning for mine valg og tolkninger. I løpet av analyseprosessen inntar jeg en mer deduktiv tilnærming ved at

jeg ser funnene i lys av eksisterende teori. Dermed beveger jeg meg mellom det induktive og det deduktive i ulike faser av prosjektet.

4.1 Min forforståelse

Forforståelsen er formet av mange ulike forhold, som oppvekst, erfaring og bakgrunn – både sosial, politisk og kulturell (Tjora, 2017, s. 25). All kunnskap er sosialt og kulturelt plassert, og det er derfor viktig at forskeren redegjør for sitt utgangspunkt. Min forståelseshorisont utgjøres av alt jeg har erfart i livet så langt. Språket har hjulpet meg med å forstå verden rundt meg og det jeg har erfart. Jeg har vokst opp i et samfunn som kan beskrives som et kunnskapssamfunn, med skoleplikt til og med 10. trinn, rett til videregående opplæring og med gode stipend- og låneordninger for høyere utdanning. Selv om skolegang på verdensbasis ikke er en selvfølge, har det alltid vært det for meg, som borger av det norske samfunnet. Det har også alltid vært en selvfølge for meg at jeg skulle utdanne meg utover videregående opplæring. Gjennom store deler av livet har jeg vært del av et vurderingssystem, eller rettere sagt av ulike vurderingssystemer. Jeg har brukt mange år av livet på skolebenken, og har alltid jobbet mye og fått føle på kroppen hvor viktig sammenhengen mellom innsats og prestasjoner er. Gode skolerresultater har gitt meg muligheter innen studier og jobb. Jeg har jobbet mange år som etterforsker og analytiker, tidvis med komplekse og alvorlige straffesaker. Gjennom arbeidet har jeg sett viktigheten av kunnskap og kompetanse, og hvordan fallgruver og feil kan utfordre rettssikkerheten. Derfor er jeg opptatt av at politistudentenes karakter i etterforskningsemnet skal være et riktig uttrykk for kompetansen den er ment å gjenspeile. Etter at jeg begynte som lærer ved PHS, har jeg blitt mer oppmerksom på mulige utfordringer ved vurderingsarbeidet. Jeg har hørt historier om studenter som er fortvilet og dypt uenige i vurderinger som er gitt, og om studenter som har blitt vurdert svært ulikt ved henholdsvis opprinnelig sensur og klagesensur. Slike historier høres også jevnlig gjennom samfunnsdebatten. Gjennom min erfaring som sensor har jeg selv vært vitne til det jeg oppfatter som (tidvis) store forskjeller i vurderingene, og hva som legges til grunn for disse. Jeg har lurt på hvordan resultatet ville blitt dersom kandidaten hadde møtt for kommisjonen jeg var en del av dagen før, eller for kommisjonen i naborommet. Ville karakteren da blitt den samme, med de tilnærmet samme begrunnelser?

Ovennevnte momenter danner bakgrunnen for mitt valg av tema for masteroppgaven. Som forsker må jeg være bevisst egen rolle og potensielle påvirkningskraft. Jeg forsker på egen

arbeidsplass og jobber som lærer innen emnet jeg samler data fra. Jeg er kollega med informantene og har selv vært sensor i emnet ved en rekke anledninger. Jeg trådte inn i forskningsfeltet med en forhåndsoppfatning – basert på egne erfaringer – om at vi vurderer ulikt. Dermed er det en fare for at jeg (kun) ser etter informasjon egnet til å styrke egne antagelser. Jeg har søkt å motvirke dette på ulike måter, blant annet gjennom grep for å unngå påvirkning i datainnsamlingen og gjennom valg av analysemetode. Jeg vil komme tilbake til det. Av ovennevnte grunner er det viktig at jeg er bevisst egen forforståelse og oppfatninger, og at jeg jobber systematisk og metodisk i alle deler av prosessen for å hente ut nyansene i empirien og for å unngå bekreftelsesfeller. Det er også viktig at jeg er transparent i min rapportering av hva som er gjort og hvorfor, slik at andre kan gjøre egne vurderinger av kvaliteten på prosjektet, herunder omkring gyldighet, pålitelighet og generaliserbarhet.

4.2 Pålitelighet, gyldighet og transparens

Pålitelighet, gyldighet og transparens er sentrale kvalitetskriterier innen kvalitativ forskning, og vil kunne påvirkes i alle prosjektets faser (Tjora s. 231-232). Jeg vil her kortfattet redegjøre for hva som ligger i hvert begrep. Hvordan jeg har gått frem for å ivareta kvalitet i prosjektet fremgår i de fortløpende redegjørelser, beskrivelser, begrunnelser og refleksjoner omkring fremgangsmåte og valg. Jeg reflekterer også over forhold som jeg anser som mulige svakheter ved prosjektet. *Pålitelighet* handler om intern logikk eller sammenheng gjennom prosjektet. For styrket pålitelighet er det viktig å fortelle om de valgene man har tatt, og å redegjøre for egen posisjon og engasjement som kan ha betydning for valgene (Tjora, 2017, s. 231, 235-238). Pålitelighet handler også om undersøkelsens resultater, og om hvorvidt andre forskere ville kommet frem til det samme resultatet (Kvale & Brinkmann, 2015, s. 137, 276). I en samfunnsvitenskapelig undersøkelse vil det ikke være mulig å gjenskape konteksten nøyaktig. En ny datainnsamling med de samme informanter vil være basert på andre forutsetninger, fordi informantene har gjennomført undersøkelsen tidligere. I tillegg vil en ny forsker ha en annen forforståelse og erfaring enn opprinnelig forsker, og vil dermed kunne prege undersøkelsen på en annen måte. Det tar meg over til kvalitetskriteriet transparens, som er nært knyttet til pålitelighet. *Transparens* handler om forskningens formidling, og i hvilken grad fremgangsmåte og valg er redegjort for og begrunnet, slik at andre kan forstå hvordan forskeren har kommet frem til sine funn (Tjora s. 232). *Gyldighet* handler om hvorvidt det er logisk sammenheng mellom prosjektets problemstilling, utforming og funn, altså om metoden undersøker det den er ment å undersøke og om forskningens funn svarer på dette. Gyldigheten

kan styrkes ved å tydelig redegjøre for hvordan problemstillingen er søkt belyst og hvilke valg som er gjort underveis (Tjora s. 232-234). Pålitelighet og gyldighet er dermed nært knyttet til hverandre.

4.3 Metodisk tilnærming

Metoden sier noe om hvordan vi bør gå frem for å utvikle kunnskap, og hjelper oss til å samle data til vår undersøkelse (Dalland, 2012, s. 111-112). Formålet med undersøkelsen danner utgangspunkt for valg av metodisk tilnærming (Halvorsen, 2008, s. 133). De metodiske valgene i prosjektet er gjort med utgangspunkt i hva jeg mener vil være mest hensiktsmessig for å belyse problemstillingen. Forskningsspørsmålene hjalp meg med å operasjonalisere problemstillingen, og til å fatte beslutninger rundt metodevalg. Jeg har benyttet spørreskjema og fokusgruppeintervju som metoder for datainnsamling. Det vil jeg komme tilbake til, men først vil jeg beskrive prosjektets oppstart.

4.3.1 Videoopptak av eksamensprestasjoner

I juni 2018, under gjennomføring av etterforskningseksamen, bestemte jeg meg noe brått for å vie masteroppgaven til temaet Vurderernes vurderinger. Forut for dette hadde temaet vært ett av flere aktuelle. For å få et realistisk grunnlag for å belyse problemstillingen, ønsket jeg å videofilme en eller to eksamensprestasjoner. En fordel med videoopptak er at det gjengir en situasjon på en detaljert og ikke-tolket måte (Tjora, 2017, s. 103). Opptakene skulle i neste omgang danne grunnlag for å samle data om vurderernes vurderinger. Selv om det ennå var måneder til masteremnets oppstart, hastet det å gjennomføre videoopptakene siden det var et år til neste gjennomføring. Jeg kontaktet veileder ved OsloMet for tanker og innspill, og jeg snakket med egen leder om prosjektet for å forankre det hos ledelsen ved PHS. Både veileder og leder stilte seg positive. Jeg utformet et informasjonsskriv som jeg sendte per epost til et utvalg studenter (vedlegg 3). Studentene ble valgt fra to klasser som skulle opp til eksamen en dag jeg selv var satt opp som reservesensor. Dermed hadde jeg mulighet til å ordne med det praktiske rundt videoopptaket den dagen. Fra de to klassene valgte jeg, ved loddtrekning, fire jenter og fire gutter fra hver klasse. Jeg håpet det var tilstrekkelig til å få samtykke fra minst en student. To studenter samtykket i at deres eksamen ble videofilmet. Jeg utformet informasjonsskriv til kommisjonsmedlemmene som skulle sensurere studentene (vedlegg 4). I skrivet ble det presisert at det kun var studenten som ville bli filmet, men at sensorenes stemmer ville bli dokumentert på lyd. Begge kommisjoner samtykket til videofilming av

prestasjonen. En kommisjon samtykket også i lydopptak av den etterfølgende karakterdiskusjonen, den andre ikke. Jeg sendte meldeskjema til NSD (vedlegg 1 og 2) hvor jeg forklarte hvorfor det hastet å gjennomføre videoopptakene, og at jeg ville sende ytterligere informasjon om prosjektet i løpet av høsten. Opptakene ble lagret på ekstern digital enhet og overført til passordbeskyttet område på min PC. Deretter ble de slettet fra den digitale enheten. Videoene ble senere lastet opp i Mediasite, på passordbeskyttet område hvor kun jeg hadde tilgang, slik at jeg kunne dele dem via link til informantene.

Jeg ser flere etiske utfordringer knyttet til innhenting og bruk av videoopptak. Opptakene kan ha utgjort en (ekstra)belastning for studenter og kommisjonsmedlemmer. Studentene var allerede i en presset situasjon. De kan ha følt på et press om å takke ja til å la seg videofilme, siden det var en faglærer som spurte. Også sensorene kan ha følt press til å si ja, og de kan ha opplevd opptaket som en belastning og stressfaktor. Riktignok var ikke kommisjonen synlig på videoen, men deres stemmer og spørsmål kunne høres og det var sannsynlig at mange informanter ville kjenne dem igjen. I verste fall kan slike forhold ha fått betydning for prestasjonen og vurderingene av den – og dermed for prosjektets gyldighet. Utarbeiding av skriv for informert samtykke ble viktig, fordi det sikret at alle parter fikk informasjon om prosjektet og hvordan opptakene ville bli brukt og behandlet.

Utover høsten fullførte jeg prosjektskissen (vedlegg 5) og sendte søknad for hele prosjektet til NSD (vedlegg 6-8). Jeg utarbeidet informasjonsskriv til mine fremtidige informanter (vedlegg 9), som også ble vedlagt søknaden, og jeg gjorde fortløpende litteratursøk- og lesning. Utarbeidelse av prosjektbeskrivelsen var tidkrevende, og innebar tenkning, grubling, refleksjoner og diskusjoner. Arbeidet viste seg å være en god investering i fortsettelsen, fordi det ga meg klarhet omkring hva jeg skulle (og ikke skulle) forske på.

4.3.2 Spørreskjema og fokusgruppeintervju

Jeg vil først begrunne hvorfor jeg valgte spørreskjema og fokusgruppeintervju som metoder for datainnsamling. Deretter vil jeg redegjøre for hvordan datainnsamlingen ble planlagt og utformet innenfor hver tilnærming, før jeg beskriver gjennomføring av datainnsamlingen og hvordan dataene ble oppbevart og behandlet. Jeg peker også på mulige styrker og svakheter som jeg har identifisert gjennom arbeidet. Utskrift av det digitale spørreskjemaet og intervjuguiden er vedlagt, se vedlegg 10-12.

Jeg valgte spørreskjema som metode for å belyse sensorenes individuelle vurderinger. Jeg ønsket å samle data fra så mange sensorer som mulig, fordi vi vet lite om hvordan vi vurderer samme prestasjon på tvers av sensorer. Min vurdering var at spørreskjema ga meg mulighet til å hente data fra langt flere enn det individuelle dybdeintervjuer ville gjort, innenfor de samme rammer. En mulig ulempe ved bruk av spørreskjema var at jeg mistet muligheten til å stille oppfølgingsspørsmål. For å imøtegå det brukte jeg åpne spørsmål med fritekstsvaret for de spørsmål som omhandlet begrunnelsen for karakterbedømmingen og bruken av kriterier og standarder. Åpne spørsmål er spørsmål uten faste svaralternativer (Tjora, 2017, s. 267). I tillegg ble sensorene, både muntlig og skriftlig, oppfordret til å være konkrete, detaljerte og utfyllende i sine svar.

I metodelitteraturen er bruk av spørreskjema ofte koblet til kvantitative metoder (se for eksempel Halvorsen, 2008, s. 141, 144). Grønmo skriver at begrepene kvantitativ og kvalitativ refererer til egenskaper ved dataene som samles inn og analyseres. Dataene er kvantitative dersom de uttrykker tall- eller mengdetermer, og kvalitative dersom de ikke uttrykkes på slik måte. Hvorvidt den samfunnsvitenskapelige tilnærmingen er kvantitativ eller kvalitativ, avhenger av om den omfatter innsamling og analyse av kvantitative eller kvalitative data (Grønmo, 2012, s. 85). Ovennevnte er i tråd med Gibbs' oppfatning om at data fra fritekstsvaret som senere skal kategoriseres for å telles, er kvantitative (Gibbs, 2018, s. 55). I min undersøkelse er dataene primært innhentet for å få frem meningsinnholdet i dem, og ikke for å telle svar eller kategorier. Det er først og fremst de åpne spørsmålene (begrunnelsene) som er egnet til å belyse problemstillingen. Spørsmålene med svaralternativer omhandler typisk forhold som lokasjons- og fagtilhørighet, og er stilt for å kunne avdekke eventuelle mønstre eller motstrid om hvordan vi vurderer som jeg ellers ikke ville ha sett. Et eksempel kan være sammenhenger mellom fagtilknytning og vurdering. Det vil være aktuelt å se noen av dataene i en form for mengdeterm, for eksempel karakterbedømmingene. Det sentrale er imidlertid ikke antall av hver karakter, men *spennet* i karakterer gitt på hver oppgave som belyser interbedømmer reliabiliteten. Av ovennevnte grunner mener jeg at jeg har gjennomført en kvalitativ undersøkelse bestående av kvalitative data.

For å belyse forskningsspørsmål nummer 2 (hvordan vektlegges de ulike vurderingene i samspillet når karakter skal settes) valgte jeg metoden fokusgruppeintervju. Fokusgrupper er en form for gruppeintervju hvor man samler et antall informanter for å diskutere ett eller flere

tema eller fokus (Wilkinson, 2004, s. 177). Metoden tilrettelegger for at man kan fange opp ulike oppfatninger rundt konkrete tema gjennom interaksjon (Tjora, 2017, s. 123). Intervjuene kjennetegnes av en ikke-styrende intervjustil hvor intervjueren setter rammer for hva som skal diskuteres, og deretter tilrettelegger for ordveksling (Kvale & Brinkmann, 2015, s. 179). I prosjektet ønsket jeg å skape en situasjon som var realistisk og gjenkjennbar sammenlignet med reelt eksamensforløp. Jeg vurderte fokusgrupper som godt egnet for dette formålet, fordi jeg kunne sette sammen grupper bestående av tre informanter fra ulike fag, som til sammen utgjorde en eksamenskommissjon. Dermed kunne informantene diskuteres seg frem til en felles kommisjonskarakter i en autentisk kontekst. Tilnærmingen ga meg også mulighet til å stille oppfølgingsspørsmål. Det var viktig fordi det var informantenes tanker og vurderinger som var av interesse, og ikke min observasjon av samspillet. Fokusgruppene var også egnet til å kaste lys over forskningsspørsmål nummer 1, idet det var sannsynlig at diskusjonene også ville omhandle de individuelle begrunnelsene.

Utforming av spørreskjema

Til forskjell fra en intervjuguide, hvor spørsmålene kan være stikkordpreget, må spørsmålene i et spørreskjema være fullstendig utformet (Tjora, 2017, s. 153). Jeg utformet spørreskjemaet i Questback. Jeg fikk egen lisens via Politihøgskolens IT-avdeling, og tilgang til passordbeskyttet digitalt område hvor kun jeg har tilgang. Av årsaker som jeg vil komme tilbake til under, delte jeg spørreskjemaet i to (vedlegg 10 og 11). Begge deler innledes med kort generell informasjon hvor informantene anmodes om å være konkrete, detaljerte og utfyllende i sine svar, siden jeg ikke har anledning til å stille oppfølgingsspørsmål. Del 1 inneholder kun ett spørsmål: «Hvilke(t) spørsmål ville du ha stilt til kandidaten, og hvorfor?». Utspørringen utgjør en tredjedel av tiden kandidaten er hos kommisjonen, og jeg ønsket å undersøke om, eller hvordan, den har betydning for vurderingen. Spørsmålet er skilt ut som en egen del fordi det, som eneste spørsmål, måtte stilles *før* informanten hadde sett hele videoen. Grunnen var at jeg antok at kjennskap til spørsmålene som ble stilt av kommisjonen (i videoen), kunne ha påvirket svarene – og dermed dataenes gyldighet. Derfor ble videoopptak av kandidatens frie redegjørelse og dette ene spørsmålet skilt ut som en egen del, kalt undersøkelsens del 1 (video del 1 og spørreskjema del 1). I tråd med en reell gjennomføring, tvang tilnærmingen informanten til å være klar med sine spørsmål umiddelbart etter den frie redegjørelsen.

Del 2 skulle distribueres til informantene umiddelbart etter at del 1 var besvart. Den bestod av

videoens siste del (utspørringen), og spørreskjema med spørsmål om vurderingen som skulle besvares i lys av hele prestasjonen. Jeg fordelte spørsmålene på tre sider, eller deler, i Questback, fordi det ga meg mulighet til å sette ønskede innstillinger for navigasjon mellom delene. På den første siden bes informantene om å opplyse sin karakterbedømming ved å huke av for et av alternativene A-F, og deretter beskrive de vurderinger og begrunnelser som lå til grunn for karakteren. Spørsmålet om begrunnelse ble stilt åpent (beskriv) for å unngå påvirkning, og er kanskje det mest sentrale for å belyse problemstillingen. Det kunne besvares med inntil 4000 ord.

Den neste siden inneholder spørsmål om bruk av pensum, sensorveiledninger, vurderingskriterier og karakterbeskrivelser i vurderingen. Først spørres informantene om de har lest pensum for henholdsvis eget og øvrige fag. Det er relevant fordi prestasjonen (av en helhetlig etterforskning) skal vurderes av sensorer fra kun tre av de fem fagene som inngår i emnet. Deretter følger tre åpne spørsmål om hvordan sensorveiledningene (fem per oppgave), vurderingskriteriene og karakterbeskrivelsene ble brukt i vurderingen. I ettertid ser jeg at jeg også kunne stilt tilsvarende spørsmål om læringsutbyttene. Imidlertid fremgår det i de eksplisitte vurderingskriteriene at læringsutbyttene er et naturlig utgangspunkt for hva som kreves. Dermed er det nærliggende å tro at de er omhandlet i begrunnelsene. Spørsmålene om pensum, sensorveiledninger, vurderingskriterier og karakterbeskrivelser er plassert *etter* spørsmålene om karakter og begrunnelse, fordi jeg i motsatt fall fryktet at de kunne påvirke begrunnelsene. Spørreskjemaet er derfor utformet slik at det ikke er mulig å navigere tilbake til forrige side etter at man først har gått videre. Det er presisert innledningsvis i skjemaet.

Den siste siden av spørreskjemaet inneholder spørsmål om informantene, herunder sensorerfaring, organisasjons-, lokasjons- og fagtilknytning og eventuelt kjennskap til kandidaten. Hensikten med spørsmålene var å samle data som – sammenholdt med karakterbegrunnelsene – kan vise eventuelle mønstre eller motstrid på tvers av eksempelvis erfaring, lokasjon eller arbeidssted. For eksempel har snittkarakteren mellom Bodø, Stavern og Oslo avveket med en hel karakter gjennom flere år. Kan dataene kaste lys over slike forskjeller? Antall informanter per fag per lokasjon er ikke mange, og jeg må derfor vurdere om det vil være mulig å si noe om slike sammenhenger uten at informanter kan identifiseres.

Intervjuguide til fokusgruppeintervju

Et fokusgruppeintervju forberedes vanligvis ved utarbeidelse av en intervjuguide. Guiden skal hjelpe med å strukturere intervjuet og vil typisk inneholde forslag til spørsmål (Kvale & Brinkmann, 2015, s. 162; Tjora, 2017, s. 153). Hvilke spørsmål som skal stilles avhenger av hva problemstillingen skal undersøke og hva som fremkommer i intervjuet. Hensikten er å forstå temaet fra informantens ståsted (Kvale & Brinkmann, 2015, s. 157). Jeg ønsket først og fremst at informantene skulle diskutere seg frem til en felles karakter fritt og uten avbrytelser fra meg. Eventuelle spørsmål ville komme til slutt og omhandle forhold ved vurderingene og vektleggingen av disse i tilfelle noe var uklart eller (for) lite utdypet. Jeg utarbeidet intervjuguiden i Excel, og la inn problemstillingen i overskriften og forskningsspørsmålet i første kolonne. De var ment som et rammeverk for øvrig innhold. Jeg la inn en kolonne for spørsmål, en for egne merknader og en for informasjon som skulle gis til informantene innledningsvis. Intervjuguiden inneholdt forslag til spørsmål som jeg antok kunne bli aktuelle, herunder hva som var grunnen til at man gikk med på å gå opp eller ned i karakter og hva som var grunnen til at noe ble vektlagt mer enn noe annet. Intervjuguiden er vedlagt (vedlegg 12).

Kvalitetssikring ved testgjennomføring

Påliteligheten i et forskningsprosjekt handler om hvorvidt det er intern logikk i forskningsprosessen, herunder sammenheng mellom problemstilling og datainnsamling (Tjora, 2017, s. 231). For å kvalitetssikre spørreskjema og intervjuguide, fikk jeg en erfaren kollega til å gjennomføre opplegget. Jeg ønsket tilbakemelding på alt fra spørsmålsstilling, spørsmålstype og relevans for problemstilling og forskningsspørsmål, og om jeg burde innta flere eller andre spørsmål. Min kollega fikk tilsendt sensormappe flere dager i forkant og ble bedt om å forberede seg som han ville gjort til en reell gjennomføring. Han fikk samme innledende informasjon som jeg planla å gi til informantene. Han gjennomførte spørreundersøkelsens to deler og sendte dem inn elektronisk, slik at jeg også fikk testet det tekniske. Jeg gjennomførte ikke fullstendig (fokusgruppe)intervju som test, men ba min kollega om å gjennomgå intervjuguiden og gi tilbakemelding på denne. Testgjennomføringen medførte noen viktige endringer i forskningsdesignet. Jeg la merke til at min kollega flere ganger trykket på pauseknappen i videoen for å notere hva kandidaten hadde sagt. Ved muntlig eksamen har man ikke slik anledning. Da prater kandidaten uavbrutt i en halvtime, og man tar de notater man rekker. For å gjøre datainnsamlingen så realistisk som mulig, valgte jeg å endre innstillingene i videoen slik at det ikke var mulig å stoppe eller pause den når den først var startet. Dermed måtte informantene foreta sin vurdering innenfor de samme rammer

som ved muntlig eksamen. Jeg la til informasjon om dette i den innledende informasjonen, slik at informantene gjorde seg utilgjengelige for forstyrrelser under videoavspillingen. Testpersonen oppfattet spørsmålene som relevante for problemstillingen. Han foreslo å legge inn spørsmål om informantenes eventuelle kjennskap til studenten på videoen, noe jeg gjorde.

4.4 Utvalg av informanter

Det er problemstillingen som danner utgangspunkt for hvilke enheter som skal inngå i en studie (Grønmo, 2004, s. 97). Innenfor mine avgrensninger var det et kriterium at sensorene var interne. Jeg måtte også ta stilling til om jeg skulle henvende meg til alle interne sensorer, uavhengig av fagtilknytning, eller om jeg skulle begrense utvalget til de fag som var representert på videoopptakene – det vil si alle fag med unntak av psykologi. Jeg vurderte begge utvalg som interessante. Ved å inkludere alle kunne jeg kanskje få dypere innsikt i vurderingen på tvers av alle fagene i etterforskningsemnet. Imidlertid vil det være slik at kommisjonssammensetningen, i større eller mindre grad, har betydning for besvarelsen i form av hva kandidaten bruker tid på. For eksempel vil en kandidat i en kriminalteknisk kommisjon typisk bruke mer tid på hvordan han vil sikre kriminaltekniske spor enn en kandidat i en psykologikommisjon, som kanskje vil vie mer tid til vitnepsykologi. Siden psykologifaget ikke var representert i kommisjonene, fryktet jeg at det kunne få negativ betydning for validiteten å inkludere dem. Derfor valgte jeg kun å henvende meg til sensorer fra fagene representert i kommisjonene i de to videoopptakene..

Jeg sendte mail til de administrativt eksamensansvarlige ved PHS' tre lokasjoner, med forespørsel om å få oversikt over interne sensorer i etterforskningsemnet. Jeg mottok en liste bestående av totalt 74 navn hvorav 21 viste seg å ikke være aktuelle, enten fordi de hadde sluttet, var tilknyttet psykologifaget eller var brukt som testperson. Jeg stod igjen med 53 interne sensorer. I kvalitative undersøkelser ønsker man å samle data fra de personene som har den beste kunnskapen om temaet som ønskes undersøkt (Halvorsen, 2008, s. 164). Etter mine vurderinger var de 53 personene best egnet til å belyse problemstillingen. Ved avgjørelsen av utvalgets størrelse må man ta både metodologiske og ressursmessige hensyn (Grønmo, 2004, s. 106). I mitt tilfelle vurderte jeg at jeg både ressursmessig og metodisk hadde kapasitet og rammer til å samle data om individuelle vurderinger (video og spørreskjema) fra alle sensorene. Imidlertid valgte jeg å begrense antall fokusgruppeintervjuer til maksimalt fire, avhengig av hvor mange som ønsket å delta, fordi jeg fryktet at

datagrunnlaget ellers ville bli for omfattende for mine rammer.

Jeg sendte mail alle 53 sensorer med kort beskrivelse av prosjektet, formål og hva deltakelse ville innebære (vedlegg 13). I mailen la jeg ved skriv om informert samtykke (vedlegg 9). Jeg ba om tilbakemelding på *om* de ønsket å delta, og i så tilfelle *hva* de ønsket å delta på: individuell vurdering (video og spørreskjema) eller individuell vurdering etterfulgt av fokusgruppeintervju (video, spørreskjema og fokusgruppeintervju). Totalt 29 takket ja til deltakelse, hvorav 19 ønsket å delta på alt mens ni ønsket å delta på den individuelle vurderingen. Jeg gjennomførte totalt fire fokusgruppeintervjuer, to per oppgave. I utvelgelsen av informanter til fokusgrupper bør forskeren sørge for at viktige karakteristika for problemstillingen er representert i utvalget. I motsatt fall kan det bli vanskelig å generalisere analytisk ut fra mønstre i empirien (Halkier, 2003; 2012, s. 137). I dette prosjektet vil jeg si at sentrale karakteristika er representert i utvalget. Fokusgruppene er satt sammen av sensorer fra de tre fagene som var representert i opprinnelig kommisjon, og av sensorer som tidsmessig gjennomførte de individuelle vurderingene samtidig. Dermed kunne de umiddelbart etter sine individuelle vurderinger settes sammen som fokusgruppe (kommisjon) på en måte som var virkelighetsnær – både i sammensetning og kontekst. Mitt kjennskap til informantene varierer. Noen kjenner jeg godt, andre kjenner jeg litt og noen kjenner jeg ikke. Under avsnittet om etiske betraktninger gjør jeg noen refleksjoner omkring dette.

4.5 Datainnsamling

Rundt en uke forut for deltakelsen mottok hver informant en epost fra meg med sensormappe og kort informasjon om hvordan datainnsamlingen ville forløpe (se vedlegg 14).

Sensormappen var identisk med den brukt ved eksamen i etterforskning våren 2018, med unntak av at jeg hadde fjernet de oppgaver og sensorveiledninger som ikke var aktuelle. Den inneholdt oppgavetekst med tilhørende sensorveiledninger, vurderingskriterier, læringsutbytter, pensumoversikt og karakterbeskrivelser. Jeg ønsket ikke å legge føringer på hvordan informantene skulle jobbe med dokumentene, og skrev derfor i eposten at de skulle forberede seg som de ville gjort til en reell gjennomføring. De fleste informanter mottok sensormappen minst en uke før deltakelsen, alternativt samme dag som avtalen ble gjort.

Dagen for datainnsamlingen møtte eller ringte jeg informant(en) cirka 10 minutter før oppstart. Jeg informerte om oppleggets to deler, og forklarte at de i del 1 ville bli spurt om

hvilke spørsmål de vil stille kandidaten og at del 2 omhandlet spørsmål om vurderingen, herunder karakterbedømming og begrunnelse. Informantene ble ikke informert om at de ville bli stilt spørsmål om bruk av vurderingskriterier, pensum, sensorveiledninger og karakterbeskrivelser, fordi jeg ikke ville risikere å påvirke innholdet i begrunnelsene. Målet var å samle valide data som var så upåvirket og «riktige» som mulig sammenlignet med hvordan informantene vanligvis vurderer. Jeg oppfordret informantene om å være ærlig og å ikke pynte på noe. Flere spurte om de skulle bedømme prestasjonen helhetlig eller i lys av eget fag. Med samme begrunnelse som over, svarte jeg at de skulle gjøre som de pleier. Deretter gikk informanten til sitt kontor, og jeg sendte vedkommende link til undersøkelsens del 1. Jeg fikk automatisk varsel via Questback når det var besvart og sendte da ny mail med link til del 2. Etter at den var innsendt var informanten ferdig med sin individuelle deltakelse. Vedkommende ble bedt om å slette epostene med linker fra meg. Jeg deaktiverte videoene i Mediasite mellom hver datainnsamling, slik at linkene ikke fungerte. Etter at siste informant hadde gjennomført undersøkelsen, slettet jeg videoene fra Mediasite. I de tilfeller hvor jeg skulle ha etterfølgende fokusgruppeintervju, møtte jeg alle tre informanter på mitt kontor forut for den individuelle vurderingen, og ga samme informasjon som beskrevet over. Vi avtalte å møtes til intervju umiddelbart etter at alle tre var ferdige med sine individuelle vurderinger.

Fokusgruppeintervjuene hadde varighet fra 21 til 43 minutter. To av intervjuene ble gjennomført på mitt kontor og to på møterom på PHS. I to av intervjuene deltok en informant via Skype, fordi vedkommende befant seg utenbys. Intervjuene ble tatt opp på lyd ved bruk av to opptaksenheter – diktafon og telefon⁴ – i tilfelle tekniske problemer eller dårlig lyd kvalitet på én av dem. Lydkvaliteten var gjennomgående god og jeg opplevde ingen tekniske problemer. Imidlertid er det klart at rammene for intervjuene som delvis foregikk via Skype var noe ulik de øvrige, blant annet ved at den kroppslige kommunikasjonen var annerledes. Tjora skriver at intervjuer gjennomført via telefon ofte får et mer formelt preg og er kortere enn intervjuer ansikt til ansikt (Tjora, 2017, s. 169). I min undersøkelse brukte vi både lyd- og bildefunksjonaliteten i Skype, slik at informantene kunne se og høre hverandre begge veier. Det synes som om kommunikasjonen fløt bra også ut fra disse rammene.

I fokusgruppeintervjuene ble informantene innledningsvis informert om at de nå var å anse som en eksamenskommissjon som skulle vurdere og karaktersette prestasjonen de nettopp

⁴ Via appen AudioMemos

hadde sett og individuelt vurdert. De ble bedt om å gjøre som de ville ha gjort i en reell setting, etter at kandidaten har forlatt rommet. For å avgrense diskusjonen til en realistisk tidsramme, anmodet jeg om at de ferdigstilte diskusjonen innen et kvarter. Imidlertid var jeg ikke veldig rigid på tidsrammen, fordi man, tross alt, har mulighet til å diskutere kandidater også i lunsjpausen eller på slutten av dagen, før sensur gis. Det sentrale for mitt formål var å få tak i hvordan de ulike vurderingene ble vektlagt i prosessen med å karakterbedømme prestasjonen. Informantene ble også fortalt at eventuelle spørsmål fra meg ville bli stilt til slutt, etter at de hadde kommet til enighet om en felles karakter. Jeg plukket ikke ut en kommisjonsleder, selv om en av sensorene ville hatt slik funksjon ved en reell gjennomføring. Grunnen til det var at jeg ønsket at alle skulle føle like stort ansvar for å føre ordet, og å bidra i diskusjonen.

Jeg ser flere mulige svakheter ved spørsmålene stilt i datainnsamlingen. Spørsmålene om bruk av vurderingskriterier, sensorveiledninger og pensum, ga ikke den utfyllende informasjonen jeg hadde håpet. Selv om de var utformet med fritekstsvaer, var beskrivelsene i stor grad kortfattet og lite informative. Eksempler på svar om bruk av vurderingskriterier var: «Jeg tittet på disse på forhånd» og «Vurderingskriteriene ga meg utfyllende og generelle føringer for vurderingen». For å forstå om, eller hvordan, disse ble brukt, måtte jeg i stor grad støtte meg til innholdet i begrunnelsene og fokusgruppeintervjuene. Spørsmålet i del 1 om hva informanten ville spurt kandidaten om, og hvorfor, ble av mange kun besvart i lys av det første spørsmålet (hva). Halvorsen skriver at man ikke bør spørre om to ting i samme spørsmål i et spørreskjema (2008, s. 145). Det har jeg gjort i dette tilfellet. Svarene kan imidlertid kaste lys over hva informantene vektlegger i spørsmålsstillingen, og hvordan dette stiller seg på tvers. I ettertid ser jeg at jeg også kunne ha inntatt spørsmål om hvordan den enkelte gikk frem for å dokumentere, eller «få med seg», innholdet i prestasjonen.

4.6 Transkribering

Transkripsjoner er oversettelser fra talespråk til skriftspråk og innebærer å gjøre den muntlige intervjusamtalen om til skriftlig tekst, slik at den blir mer tilgjengelig for analyse. I praksis innebærer transkripsjon å abstrahere og dekontekstualisere det som i realiteten var en sosial samtale mellom mennesker i samspill (Kvale & Brinkmann, 2015, s. 204-205). Man vil alltid miste noe på veien fra muntlig til skriftlig form, for eksempel visuelle ledetråder som kroppsspråk og informasjon om stemning (Tjora, 2017, s. 175). For å best mulig huske

konteksten for uttalelsene, prioriterte jeg å transkribere intervjuene så raskt som mulig, senest innen en uke etter at de ble gjennomført. Tjora skriver at det å lese en transkribert tekst fra et intervju man selv har deltatt i, vil kunne ta en tilbake til intervjusituasjonen (Tjora, 2017, s. 175). Det er i overensstemmelse med mine opplevelser. I transkripsjonene ble hver informant omhandlet med benevnelsen for det faget vedkommende representerte, slik at ingen personidentifiserende opplysninger ble nedtegnet.

I tråd med Tjoras anbefaling (2017, s. 173) valgte jeg fullstendig transkribering av intervjuene. Det var viktig å få tak i meningsinnholdet i uttalelsene, slik at jeg dannet godt grunnlag for senere analyse. Derfor nedtegnet jeg som hovedregel nøyaktig det som ble sagt, slik det ble sagt. Imidlertid hendte det at jeg utelot tilleggsord eller uttrykk, som «eh» eller «på en måte», men bare i tilfeller der jeg vurderte at disse var uten betydning for meningsinnholdet. Det hendte også at jeg endret på ordenes rekkefølge i en setning for å gjøre dem mer leservennlige. Informasjon om tenkepauser, kroppsspråk og tonefall ble ikke nedtegnet med mindre jeg oppfattet at de hadde betydning for meningsinnholdet. Jeg tilstrebet å være tro mot kilden til informasjonen, slik at det var den som dannet grunnlag for senere analyse – og ikke mine tolkninger av hva informanten kan ha ment. Mine mulige tolkninger ble holdt adskilt i eget merknadsfelt. Imidlertid vil det ikke være mulig å nedtegne det noen har sagt, uten å være i en fortolkningsprosess. Bare det å sette punktum og komma innebærer tolkning (Kvale & Brinkmann, 2015, s. 212). Det samme gjør vurderingene knyttet til endring på rekkefølgen og fjerning av enkeltord. Dermed risikerte jeg å endre meningsinnholdet, og dermed påliteligheten i dataene, i transkriberingsprosessen. Jeg kvalitetssikret arbeidet ved å høre gjennom alle lydfiler på nytt, og sammenholde dem med tekstfilene. Ideelt sett burde jeg ha vist transkripsjonene til informantene som deltok, og fått deres vurderinger av den skriftlige nedtegningen. Det ble ikke gjort. Den ekstra gjennomgangen medførte ingen større endringer. All transkribering ble gjort i NVivo, som har funksjonalitet for sammenkobling av lyd og tekst. Kvaliteten på lydopptakene var gode.

Under transkriberingen hørte jeg tilfeller hvor jeg stilte spørsmål på lite hensiktsmessige måter. Noen spørsmål var lukket og ledende. Andre inneholdt flere spørsmål i ett. Jeg identifiserte tilfeller hvor jeg ikke fulgte opp det informanten sa i tilstrekkelig grad, men opplevde tidvis at andre informanter i samme intervju tok opp tråden der jeg burde ha stilt spørsmål, og at sentrale tema derfor likevel ble diskutert og dekket. Gjennomgående opplevde jeg at det ikke var behov for mange spørsmål, fordi informantene uttrykte seg rikt og detaljert

gjennom den innledende karakterdiskusjonen og vurdering av prestasjonen.

Informasjonsbehandling

Datainnsamlingen foregikk over en periode på to måneder. Jeg gjorde fortløpende avtaler, og holdt oversikt over hvem og når i et excelark. I oversikten ble hver informant gitt et unikt nummer. Nummerets kobling til informantens navn ble ført i en egen liste (koblingsnøkkel), og lagret på adskilt, passordbeskyttet område hvor kun jeg hadde tilgang. På samme plassering ble lydfilene fra fokusgruppeintervjuene lagret, siden de kunne være personidentifiserende gjennom stemme. Lydfilene ble slettet fra de to opptaksenhetene. All epost-korrespondanse med informantene ble slettet i etterkant av datainnsamlingen. Gjennom prosjektet førte jeg fortløpende logg i NVivo. Jeg hadde egne logger for metode, analyse og teori, og en for alt annet. Loggene var nyttige for å «ta vare på» oppdøkkende tanker. Etter hvert intervju skrev jeg ned egne refleksjoner, og ettersom intervjuene forløp så jeg mønstre og motstrid i dataene. Tilsvarende oppdagelser gjorde jeg ved transkribering og nedbryting.

4.7 Etiske betraktninger

Gjennom arbeidet med et forskningsprosjekt vil forskeren stå overfor en rekke etiske problemstillinger. De nasjonale forskningsetiske komiteene har utarbeidet og publisert generelle og fagspesifikke forskningsetiske retningslinjer. De er ment som verktøy for å fremme god og ansvarlig forskning gjennom å konkretisere forskersamfunnets grunnleggende normer og verdier (FEK, 2016; NESH, 2016, s. 4-5). Overordnet regulerer retningslinjene normer for god vitenskapelig praksis for kunnskapsutvikling (åpenhet, etterprøvbarehet), normer for forskersamfunnet (redelighet, etterrettelighet), forpliktelser til personene som deltar i forskningen (konfidensialitet, informert samtykke) og forskningens relasjon til resten av samfunnet (uavhengighet, forskningsformidling) (NESH, 2016, s. 6). Mine refleksjoner over etiske aspekter ved prosjektet er i hovedsak beskrevet fortløpende der de hører hjemme. Blant annet har jeg beskrevet ulike grep for å ivareta informantenes anonymitet og integritet (i datainnsamling, informasjonsbehandling og formidling), bruk av informert samtykke for alle involverte, søknad til, og godkjenning fra, NSD, grundig redegjørelse av fremgangsmåte og valg i forskningsformidlingen, og god henvisningsskikk. Jeg har også tilstrebet å ha et bevisst forhold til egen forforståelse og kunnskapssyn, og å synliggjøre hvordan disse kan ha påvirket mine valg og fokus.

Et sentralt forskningsetisk aspekt er at informantene ikke skal komme til skade. Derfor er det viktig å reflektere over eventuelle ubehag som fremgangsmåten kan utsette dem for (Tjora, 2017, s. 175). Under avsnittet om videoopptak reflekterte jeg derfor rundt mulige etiske utfordringer knyttet til innhenting og bruk av videoopptak – både for sensorene og studenten som ble filmet. Det kan også tenkes at informantene opplevde min dobbeltrolle som forsker og kollega (og medsensor) som en ulempe eller skade. For eksempel kan de ha følt seg kontrollert, eller følt press til å prestere, siden jeg har inngående kjennskap til både eksamensform og det faglige innholdet i emnet. Slike forhold kan ha medført at informantene besvarte spørsmålene i datainnsamlingen annerledes enn de ellers ville ha gjort, og dermed fått betydning for prosjektets gyldighet. Deltakerne kan også ha følt press til å delta som følge av min kollegiale relasjon til dem. Tjora skriver at selv om informanter er opplyst om sitt informerte samtykke og retten til å trekke seg, kan de likevel vegre seg for å «svikte» (Tjora, 2017, s. 48). Jeg har forsøkt å minimere slike konsekvenser, blant annet gjennom tydelig kommunikasjon omkring prosjektets formål, viktigheten av «riktige» data og at ingen vil bli omhandlet med navn i forskningsrapporten. Jeg har også vært tydelig, både skriftlig og muntlig, på at deltakelse er frivillig. I rapporteringen av forskningen har jeg også pekt på forhold som jeg anser som mulige svakheter, fordi jeg ønsker at leseren skal kjenne til forhold egnet til å påvirke resultatet av undersøkelsen. Gjennom hele prosessen har det vært et sentralt prinsipp at ingen informanter skal kunne identifiseres, verken gjennom informasjonsbehandling eller rapportering – i tråd med sentrale etiske aspekter som konfidensialitet, respekt og tillit, som gjelder i all forskning (Tjora, 2017, s. 46).

Jeg har nå redegjort for min forforståelse og for hvordan jeg har gått frem for å belyse problemstillingen. I det neste kapitlet vil jeg beskrive analysen og undersøkelsens funn.

5. Analyse og funn

I det følgende vil jeg gjøre rede for hvordan jeg har kommet frem til mine funn, og hva disse er. Som nevnt innledningsvis har jeg valgt å omhandle analyse og funn samme kapittel, som følge av den nære sammenhengen mellom dem.

5.1 Analyseprosessen

Analyse er et granskningsarbeid der utfordringen ligger i å finne ut hva materialet har å fortelle (Dalland, 2012). Analysen starter idet forskeren trer inn på forskningsfeltet, og strekker seg gjennom hele prosessen (Postholm, 2010, s. 99). Jeg ønsket å starte induktivt, med empirinære koder, og etterhvert bevege meg i retning av det deduktive ved å koble inn teori. Ideelt sett skulle jeg gjennomgått empirien helt forutsetningsløst, men det er ikke mulig. Allerede tidlig leste jeg teori og forskning om temaet som fikk betydning for hva jeg samlet data om. Spørsmålene om sensors erfaring og bruk av vurderingskriterier er eksempler på det. Også min egen forforståelse har påvirket mine fokus og valg underveis i prosjektet. Selv om det ikke er mulig å være helt induktiv, kan jeg likevel kode åpent. Gibbs beskriver en empiridreven tilnærming til analyse ved bruk av åpen koding, hvor man starter uten forhåndsdefinerte kategorier. Han kaller det tematisk koding og kategorisering (2006, s. 53, 60). Målet er å forstå hva empirien kan fortelle om det en undersøker og – så langt det lar seg gjøre – unngå å basere tolkningen på eksisterende teori og egen forutinntatthet. Jeg vurderte at Gibbs' empirinære tilnærming passet godt for mitt prosjekt og formål. Det tar høyde for, og anerkjenner, at det ikke er mulig å jobbe forutsetningsløst, men søker å motvirke forutinntatthet gjennom systematisk og metodisk tilnærming til behandling av dataene. Derfor valgte jeg åpen koding, inspirert av Gibbs, som analysetilnærming. I tråd med Gibbs anbefaling om at koding skjer best ved bruk av et digitalt analysesystem, (2018, s. 55, 56), valgte jeg å bruke NVivo til dette arbeidet.

Selve koden, eller kategorien, er en kortfattet benevnelse som representerer tekstutdrag fra datagrunnlaget og som eksemplifiserer en tematisk ide. Kodene bør være så teoretiske og analytiske som mulig, heller enn deskriptive. Det kan hjelpe oss med å forstå dataene på måter som strekker seg utover informantens uttalelser og som blir mer generelle (Gibbs, 2018, s. 54, 73). Gibbs gir følgende eksempel: I stedet for «joining a sportsclub», som er deskriptiv, kan man bruke koden «Activity to make friends» eller «Commitment to keeping fit», som mer generelt beskriver hva uttalelsen handler om (Gibbs, 2018, s. 64). Slik koding vil, ifølge Gibbs, gi en mer overordnet oversikt over hva dataene handler om og gjøre det lettere å sammenstille dem på tvers av kilder. Begrepene *kode* og *kategori* blir her brukt om det samme, i tråd med forståelsen beskrevet i dette avsnittet.

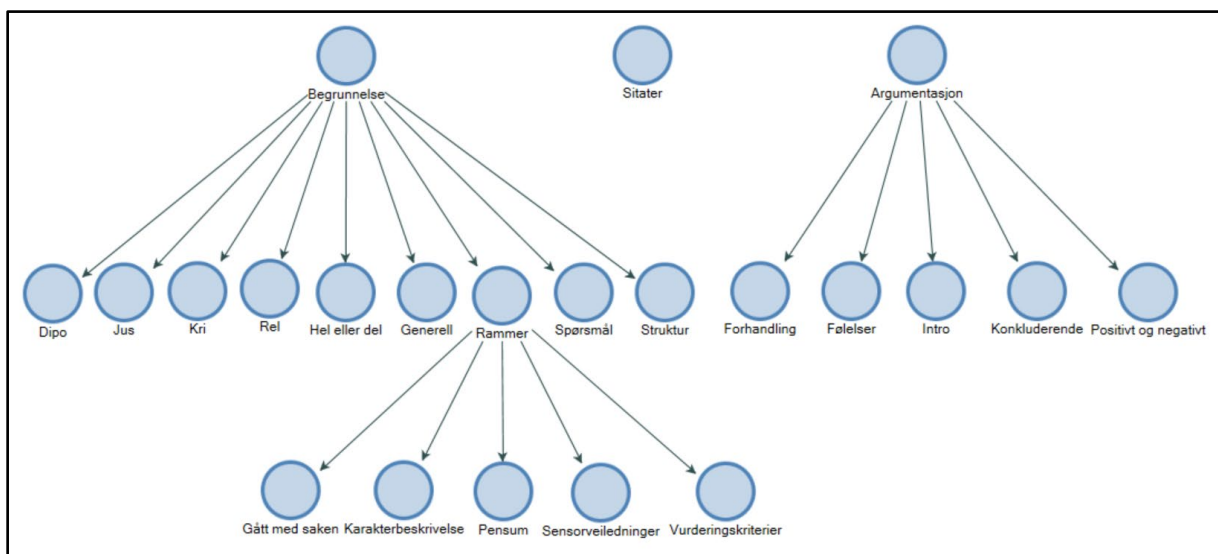
Jeg startet med fokusgruppeintervjuene. Hver setning ble lest og vurdert for koding – enten som enkeltstående uttalelse eller som en del av den sammenhengen den var uttalt i. Gibbs

omtaler det som «line-by-line coding», et begrep hentet fra Grounded Theory (Charmaz, 2014; Gibbs, 2018, s. 73). For å motvirke faren for at uttalelsen skulle miste sin betydning ved å bli tatt ut av kontekst, tok jeg med så mye jeg mente var nødvendig for å bevare meningsinnholdet. Jeg støttet meg til Gibbs' prinsipper om å tenke overordnet, og laget kategorier som uttrykte hva uttalelsene handlet om, heller enn konkret innhold. Eksempelvis opprettet jeg kategorien *Struktur*, som jeg definerte at skulle omfatte begrunnelser rettet mot *formen* på fremføringen (ikke-faglige forhold) – for eksempel klesdrakt, kandidatens fremtoning eller struktur i fremføringen. Kategorien ble opprettet fordi jeg så at mine data inneholdt flere uttalelser som belyste slike forhold. I NVivo opprettes automatisk link mellom tekstutdrag og kildefil ved koding, og jeg hadde derfor hele tiden kontroll på referansen til hver kodet uttalelse. Ved endt nedbryting kunne jeg hente ut alle tekstutdrag som var knyttet til samme kategori. Ifølge Gibbs er dette en av de store fordelene ved tematisk analyse, ettersom den muliggjør uthenting av de tekstutdrag fra empirien som omhandler samme tema – på tvers av de mange kildene til informasjon (Gibbs, 2018, s. 54, 73). Siden tekstutdragene beholdes i sin opprinnelige form ved nedbryting i NVivo, legges det til rette for at analysen baseres på kildens uttalelser, og ikke mine tolkninger av dem i nedbrytingsfasen.

Ifølge Gibbs er det viktig å definere kodene for å beskrive den analytiske ideen den referer til, og for å sikre at kodingen er reliabel ved at den gjennomføres på en systematisk og konsekvent måte (s. 56-57). Jeg definerte alle koder og nedtegnet definisjonene i NVivo. Gibbs anbefaler bruk av slik kodebok for å holde oversikt over den til enhver tid gjeldende kodestruktur, dens hierarkiske organisering og tilhørende definisjoner (2018, s. 55). Etter første runde med nedbryting av fokusgruppeintervjuene, stod jeg igjen med 17 kategorier. Dataene fra de åpne spørsmålene i spørreskjemaet ble brutt ned etter samme systematikk. Dette arbeidet genererte ingen nye kategorier. Det er ikke overraskende, fordi mange av de samme vurderingene og begrunnelsene gikk igjen begge steder. Hadde jeg derimot startet med spørreskjemaene, ville jeg trolig sett behov for nye kategorier for fokusgruppene, fordi disse også genererte kategorier som favnet argumentasjon, overtalelse og begrunnelser i samspillet med andre. Spørsmålene med svaralternativer krevde betydelig mindre bearbeiding enn de åpne, fordi de allerede var ferdig strukturerte i Questback. Jeg importerte dem til NVivo, slik at de kunne sammenholdes med øvrig empiri.

Det ble etterhvert tydelig at jeg hadde vært for detaljert i kodingen, blant annet fordi jeg stadig måtte knytte tekstutdragene til flere kategorier som på sett og vis handlet om det samme.

Derfor valgte jeg å slå sammen koder til bredere kategorier, som eventuelt kunne brytes ytterligere ned ved behov. For eksempel hadde jeg en kategori kalt *Helhet* og en kalt *Faginndelt*, hvor førstnevnte favnet uttalelser som var helhetlig orientert, og sistnevnte uttalelser som vektla fagspesifikke elementer. Ved å slå disse sammen til én kategori (hel eller del), ville jeg kode begge vinklinger og eventuelle kombinasjoner av disse, men nå under en og samme kategori. Jeg gjorde flere tilsvarende sammenslåinger av koder som jeg vurderte at hørte sammen. I tillegg laget jeg et hierarkisk system med hovedkategorier og underkategorier, slik at noen kategorier som hadde felles innbyrdes tema ble plassert under en større hovedkategori. Jeg stod igjen med følgende kodesystem:

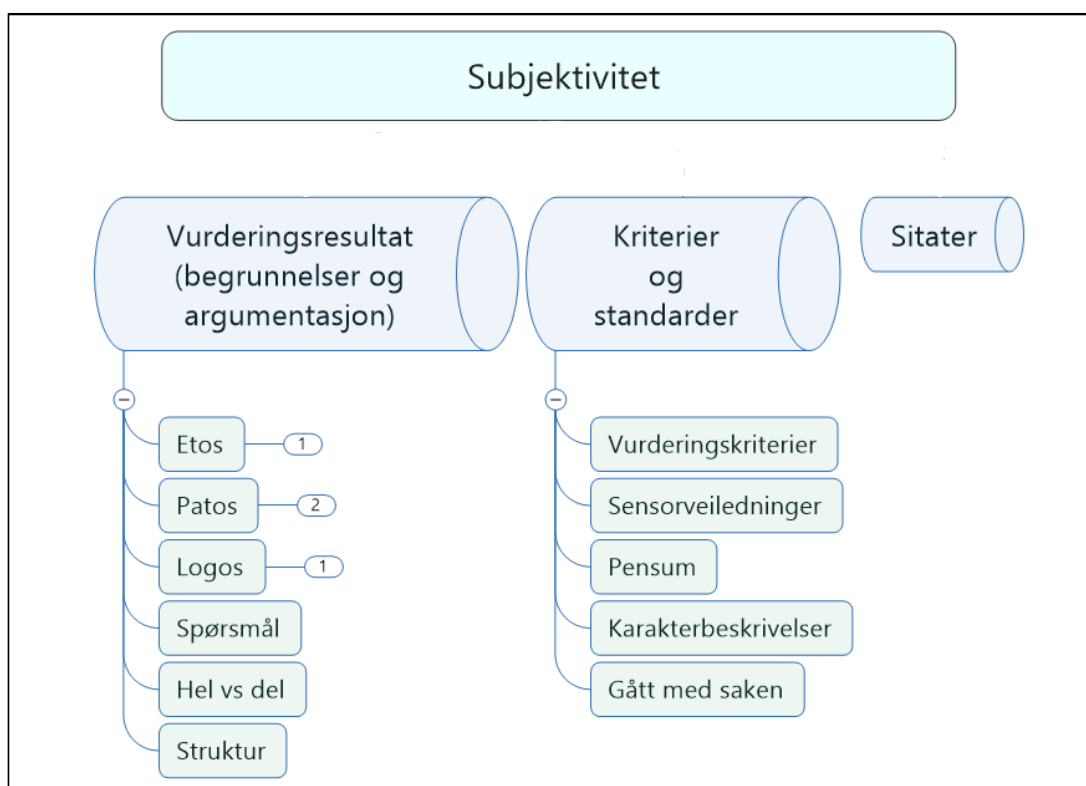


Figur 3: Kodesystem etter nedbryting trinn 2 – empirisk orientert

Fra empiriske til teoretiske kategorier

Parallelt med nedbrytingen leste jeg mye teori, og det ble mer og mer tydelig at det ga mening å se empirien i lys av de teoretiske perspektivene. Derfor valgte jeg å (om)strukturere empirien slik at den ble mer teoretisk rettet. Mange kategorier ble beholdt, men jeg organiserte dem noe annerledes. Som tidligere ble hver kategori definert av meg, slik at jeg hadde en klar oppfatning av hvilken type data som hørte til hvor (se vedlegg 15). Jeg tok utgangspunkt i begrepene slik de ble beskrevet i teorien, men så behov for noen tilpasninger. Fra det femdimensjonale rammeverket opprettet jeg kategorier av dimensjonene *Kriterier og standarder* og *Vurderingsresultat*. Det sentrale var hvordan eksamensprestasjonen ble vurdert gjennom bruk av kriterier, begrunnelser og argumentasjon. *Kriterier og standarder* ble brukt om de rammer som ble lagt til grunn ved karaktervurderingen. Det kunne være eksplisitte

kriterier eller standarder, eller andre kriterier som informantene støttet seg til ved vurderingen. *Vurderingsresultat* ble brukt om karakteristika ved begrunnelsene og argumentasjonen som dannet grunnlag for vurderingen. De øvrige tre dimensjoner (*oppgaven, fysisk kontekst og sosial kontekst*) er ikke direkte omhandlet i empirien, men utgjør et viktig bakteppe som kontekst. Derfor er de en del av valgt teori, men ikke egnet som analysekategorier for mitt formål. Jeg opprettet også analysekategorier for overtalelsesmidlene *etos, patos og logos*, fordi det kunne bidra til å øke min forståelse omkring hva som ble vektlagt i samspillet når karakter skulle settes. Kategoriene hjalp meg med å sortere ulike måter å argumentere på som jeg hadde oppdaget gjennom den åpne kodingen. Til slutt opprettet jeg en kategori kalt *subjektivitet*, men erfarte raskt at den var uegnet som analysekategori i betydningen av å favne uttalelser som belyste subjektiviteten i vurderingene. Grunnen til det var at nærmest alle dataene er preget av subjektivitet, herunder begrunnelser, spørsmål, karaktersetting og bruk av kriterier og standarder. Derfor vurderte jeg at subjektiviteten kunne plasseres som en overordnet kategori over alle de analytiske kodene, og på den måten visualisere et sentralt funn. Etter å ha omstrukturert dataene i tråd med de teoretisk orienterte kategoriene, stod jeg igjen med følgende kategorier og underkategorier⁵:



Figur 4: Kodesystem etter nedbryting trinn 3 – teoretisk orientert

⁵ Her vises to ledd. For noen av kategoriene hadde jeg ytterligere underkategorier, hentet fra den første nedbrytingsprosessen.

Validitet i analyse er knyttet til hvor godt kategoriseringen representerer informantenes erfaringer. For å oppnå god validitet er det viktig med en gradvis utvikling av kategoriene (Postholm, 2010, s. 170). Mine kategorier har, slik jeg ser det, blitt utviklet og videreutviklet gjennom en rekke prosesser. Fokusgruppeintervjuene ble tatt opp på lyd og deretter transkribert. Dataene fra spørreskjemaet ble lest fortløpende og ført i en samlefil i Excel. Disse prosessene ga meg tidlig tanker om mulige koder. Deretter har jeg – gjennom nedbryting og strukturering av dataene inspirert av Gibbs' metode – gjennomgått dataene i flere runder for å komme frem til ovennevnte kategorier. Etter endt nedbryting hentet jeg ut oversikt over karakterbedømmingene, både de individuelle og kommisjonskarakterene.

Jeg gjorde flere stikkprøver av mine tolkninger for å teste påliteligheten av dem. Jeg fikk hjelp av en kollega til å se gjennom deler av materialet (anonymisert) og å uttrykke sine vurderinger. For eksempel leste vedkommende en begrunnelse som jeg på forhånd hadde brutt ned, kategorisert og knyttet egne kommentarer til (som ikke var synlig for min kollega). Våre vurderinger samsvarte, ved at begge mente begrunnelsen var fagfokusert og lite helhetlig, og lite utdypet. Min kollega gjennomgikk flere uttalelser på tilsvarende måte, men langt fra alle. Gjennomgangen viste ingen store ulikheter i våre vurderinger.

Jeg har nå redegjort for hvordan jeg har behandlet dataene for å øke egen forståelse for hva de kan fortelle. I det følgende vil jeg beskrive mine funn.

5.2 Funn

Funnene skal belyse problemstillingen og forskningsspørsmålene, og jeg gjentar derfor disse her:

Hvordan vurderer interne sensorer ved Politihøgskolen muntlige eksamensprestasjoner?

1. Hvilke begrunnelser legges til grunn for karakterbedømmingene?
2. Hvordan vektlegges de ulike vurderingene i samspillet når karakter skal settes?

Først vil jeg kortfattet redegjøre for det jeg anser som hovedfunn, og deretter utdype bakgrunnen for disse med utgangspunkt i strukturen fra analyseprosessen, med tilhørende analysekategorier. De omhandlede sitater er valgt ut fordi jeg mener de er godt egnet til å illustrere funnene.

5.2.1 Oppsummering av hovedfunn

Mine funn viser at vurdererne vurderer ulikt – i karakterbedømming, begrunnelser, spørsmålsstilling og i bruken av kriterier. Sensorene verdsetter og vektlegger delvis ulike momenter ved samme prestasjon – både det som trekker opp og det som trekker ned. Karakterene varierer med 2 - 2,5 karakterspenn på samme oppgave, og ulikhetene er nesten like store på tvers av sensorer innen samme fag som på tvers av fagene. De erfarne sensorene synes å være noe strengere enn de mindre erfarne.

Vurderingene gjøres som oftest med utgangspunkt i det faget sensor representerer. Flere informanter begrunner sitt individuelle fagfokus med at helheten vil ivaretas i kommisjonene, men mine funn kan tyde på at det ikke nødvendigvis stemmer. Kommisjonskarakteren synes å bli satt et sted midt imellom kommisjonsmedlemmenes individuelle vurderinger, og de to fagene som ikke er representert blir i liten grad hensyntatt i vurderinger og begrunnelser. Dermed kan det synes som om det, i alle fall tidvis, er et gap mellom det studentene er bedt om å besvare (en helhetlig etterforskning) og det som vektlegges i vurderingene.

Bruken av kriterier og standarder varierer. Det samme gjør oppfatninger av hvordan, eller om, disse skal brukes. Noen har en analytisk tilnærming til vurderingen, og opplyser å bruke de eksplisitte vurderingskriteriene aktivt. Andre har en mer holistisk tilnærming, og støtter seg til implisitte kriterier basert på eget skjønn og erfaring. Flere opplever de eksplisitte kriteriene som et hinder for en helhetlig vurdering. Informantene som opplyste at de ikke brukte de eksplisitte kriteriene var alle erfarne sensorer i etterforskningsemnet med minst fire års erfaring. De fleste informanter støttet seg til de kvalitative karakterbeskrivelsene ved bedømmingen, men det varierte hvilke kriterier i beskrivelsene som ble vektlagt.

En del av de begrunnelser som ligger til grunn for karakterbedømmingen, er preget av å være lite utdypet. Det gjør det tidvis vanskelig å få tak i hvilke konkrete styrker og svakheter ved prestasjonen som dannet grunnlag for karakteren. Forhold ved kommunikasjonen – både i argumentasjon, innhold og rekkefølge – synes å ha betydning i samspillet når karakter skal settes. For eksempel fant jeg at det første som blir sagt i karakterdiskusjonen legger føringer både for de påfølgende uttalelser fra øvrige sensorer, og mer generelt for fokuset i karakterdiskusjonen. En innledning hvor det snakkes negativt om prestasjonen synes å dra med seg et mer negativt fokus gjennom diskusjonen, enn en innledning som fremhever styrker og svakheter ved prestasjonen. Diskusjonene i kommisjonene karakteriseres ikke bare av

faglige argumenter, men også av argumenter knyttet til forhold ved sensorene, for eksempel deres erfaring, og argumenter egnet til å vekke følelser. Bastant og konkluderende argumentasjon synes å ha større gjennomslagskraft enn argumenter av mer undrende eller spørrende art.

5.2.2 Vurdererne vurderer ulikt – i karakterer, begrunnelser og spørsmålsstilling

De individuelle karakterene varierer med inntil 2,5 karakterer. Eksamensoppgave 1 ble vurdert i karakterspennet mellom A til C/D, og eksamensoppgave 2 i karakterspennet C/D til F. Funnene viser at de uerfarne sensorene står for hovedvekten av de «snilleste» karakterene innenfor hver oppgave. Med uerfarne mener jeg her informanter som har sensurert ved denne eksamen kortere enn fire år. I underkant av halvparten av informantene faller inn under denne kategorien. Under vises de individuelle karakterene som de to eksamensprestasjonene ble vurdert til. Oversikten viser også vurderingene per fagtilknytning, men selve faget er anonymisert siden jeg bare har fire til seks informanter fra hvert fag per oppgave, og ikke vil risikere at noen skal kunne identifiseres.

Eksamensbesvarelse	Vurdererens fagtilknytning	Karakter	Eksamensbesvarelse	Vurdererens fagtilknytning	Karakter
1	Fag 1	A	2	Fag 1	D
		B			F
		C			F
		C			F
	Fag 2	A		Fag 2	D
		B			D
		C/B			D
		C			D
		C/D			E
	Fag 3	A		Fag 3	C/D
		B			D
		B			D
		C			D
		C			E
		C			E

Figur 5: Individuelle karakterbedømminger med tilhørende fagtilknytning

Oversikten viser at karakterbedømmingen innad i samme fag varierer i spennet mellom en og 2,5 karakterer. For fag 2 i oppgave 2 vises høy interbedømmer reliabilitet, ved at prestasjonen ble vurdert tilnærmet likt (4xD og 1xE). Felles for vurderingene er at de tar utgangspunkt i det aktuelle faget, noen mer helhetlig orientert enn andre. Imidlertid legger disse sensorene delvis vidt forskjellige premisser til grunn for karakterslutningene, både gjennom spørsmålsstilling og i begrunnelser. Dermed er sensorsamsvaret lavere enn det tilsynelatende kan fremstå

gjennom karaktersettingen. Ser man på interbedømmer reliabiliteten for det samme faget i eksamensbesvarelse 1, er den betydelig lavere og varierer i spennet mellom A og C/D. Innenfor de øvrige fagene, varierer karaktersettingen i spennet mellom 1,5 og 2,5 karakterer.

Karakterene varierer mindre på tvers av kommisjonene enn på tvers av individer, men mitt sammenligningsgrunnlag for kommisjoner er også betydelig mindre enn for de individuelle vurderingene. For eksamensoppgave 1 skilte én karakter mellom de to kommisjonene (A og B), mens det for eksamensoppgave 2 ikke var avvik (E og E). I den opprinnelige kommisjonen for eksamensoppgave 2, uttrykte alle tre sensorer gjennom karakterdiskusjonen at de hadde vurdert prestasjonen til en D. De ga også D som samlet kommisjonskarakter.⁶ Alle tre sensorer i den opprinnelige kommisjonen var enige om at de ønsket å høre neste kandidat, som skulle besvare samme oppgave, før de bestemte endelig karakter – i tråd med normbasert vurdering. I figur 6 vises karaktervurderingen til hver kommisjon, samt de individuelle karakterene som hvert kommisjonsmedlem i forkant hadde bedømt prestasjonen til. Jeg har ikke inntatt de individuelle karakterene til den opprinnelige kommisjonen, fordi de ikke ble samlet inn individuelt i forkant – som for øvrige informanter.

Eksamensbesvarelse	Kommisjon	Individuell karakter	Kommisjonskarakter
1	Opprinnelig		B
	1	B A A	A
	2	A B C	B
2	Opprinnelig		D
	3	D F E	E
	4	D F D	E

Figur 6: Kommisjonskarakterer for hver fokusgruppe og for den opprinnelige kommisjonen som vurderte kandidat 1, samt kommisjonsmedlemmenes individuelle karakterbedømminger

⁶ Den opprinnelige kommisjonen for eksamensoppgave 1 samtykket ikke i opptak og bruk av deres karakterdiskusjon (kun karakteren), og er derfor ikke omhandlet.

Kommisjonskarakterene synes å havne midt imellom de individuelle vurderingene. I en av fokusgruppene gikk alle informantene bort fra sin opprinnelige karakter, ved at D, D og F ble til en E. På spørsmål om hvorfor informantene var villige til å fravike sin opprinnelige karakter, var det samstemthet om at de individuelle vurderingene var gjort med hovedfokus på eget fag, mens kommisjonskarakteren gjenspeilet en mer helhetlig vurdering. Følgende sitat kan illustrere det: «Vi møttes jo med utgangspunkt i eget fag, tenker jeg, og så endte vi på karakteren som helhet». Generelt fant jeg at de individuelle vurderingene og begrunnelsene i stor grad tok utgangspunkt i faget som sensor representerte, noen mer detaljfokuserte enn andre på rent fagspesifikke forhold. Flere informanter begrunnet sitt individuelle fagfokus med at helheten ville bli ivaretatt i kommisjonene. Under vises et utdrag hentet fra diskusjonen i en av fokusgruppene, som illustrerer oppfatninger av forholdet mellom fagfokus og helhetsvurdering:

Fag 1: Sånn etterforskningsmessig har jeg ikke noen problemer med å kanskje gå med på en E. men jeg er ikke oppe på en D. jeg er langt ifra en D

Fag 2: Det er jeg og, langt unna

Fag 3: Neimen jeg, jeg var isolert på en D i mitt fag fordi vi gjorde det på forhånd, før vi ble enige. Jeg vet ikke om jeg gjorde feil der, men jeg satte D ut fra eget fag isolert sett ... Men samlet, som vi skal, så mener jeg at kandidaten er på en E.

Sitatet illustrerer også et annet funn, nemlig at karakteren i stor grad ble basert på de tre fagene som var representert i kommisjonen, og ikke alle fem. Det ble tydelig gjennom karakterdiskusjonene. Imidlertid er det ikke slik at momenter fra de andre fagene var fullstendig fraværende, eller at de overhodet ikke ble nevnt. I en kommisjonene (fokusgruppene) ble begge fag nevnt, men uten at det konkret ble snakket om hvordan momenter fra disse fagene ble (eller ikke ble) ivaretatt i besvarelsen – og uten at slike forhold syntes å ha betydning for den endelige vurderingen. I en annen kommisjon ble ett av de to fagene hensyntatt, ved at man så det som et minus at faget var lite omhandlet i prestasjonen. Følgende sitater kan illustrere dette funnet: «Kandidaten har 30 minutter i tre emner. Det er en helhetsvurdering» og «Det skal være en samlet vurdering. Alle tre fagene skal settes opp mot to kriterier – selvstendighet og forståelse» og «Vi har en svak D, en svak E og en F. Da er spørsmålet hvor vi ender».

Begrunnelser og spørsmålsstilling:

Hva som legges til grunn for karakteren og hvilke spørsmål informantene ville stilt til kandidaten, varierer. Det er også ulike vurderinger av hvorvidt spørsmålsrunden hadde betydning for karakteren. Jeg fant ikke noe entydig mønster for verken begrunnelser eller spørsmål, heller ikke innad i samme fag. Noen har et overordnet fokus, andre er opptatt av detaljer. Noen begrunner utfyllende og konkret, andre mer kortfattet og generelt. Noen fokuserer (mest) på svakheter og mangler, andre på styrker. Noen fokuserer på eget fag, andre er mer helhetsorientert. Noen vektlegger hva kandidaten kunne eller ikke kunne svare på i spørsmålene, andre ikke. Noen vektlegger forhold som struktur, systematikk, (oppfatning av) forberedelser og disponering av tid, andre ikke. Hvilke helhetsmomenter eller detaljer det fokuseres på, enten de er fagspesifikke eller overordnet, varierer også. Dette betyr imidlertid ikke at ikke mange av de samme tema, eller begreper, går igjen. For eksempel nevnes momenter som bevistema, hypoteser, IP-adresser, tvangsmidler, målstyrt etterforskning, objektivitet, sporsikring og etterforskningsprosess av mange informanter, men om, og hvordan, disse brukes, vektlegges og begrunnes, varierer. Under har jeg valgt ut to begrunnelser fra samme fag og samme oppgave, som illustrerer noe av ulikhetene og variasjonene i begrunnelsene. Den siste begrunnelsen er kun delvis gjengitt, fordi den ellers ville ha fylt en hel side.

Begrunnelse 1:

Kandidaten viser stor grad av selvstendighet og forståelse for hele etterforskningsfaget. Kandidaten har god forståelse for jussen, vet hva de ulike bestemmelsene og deres vilkår innebærer. Jeg kunne med fordel "trukket litt ned" for noe svak forholdsmessighet (nevner for eksempel ikke sakens alvorlighetsgrad som et moment).

Begrunnelse 2:

Kandidaten tar utgangspunkt i sentrale bestemmelser i straffeprosessloven § 224, § 225 og § 226 a og b. Videre knytter kandidaten formålsbestemmelsen til at de objektive og subjektive straffbarhetsvilkår må være oppfylt. Kandidaten glemte vel i farten å nevne at det ikke må foreligge straffrihetsgrunner. Kandidaten redegjør for objektivitetsprinsippet og viktigheten av å unngå justisfeil i etterforskningen hvilket er bra at det er fokus på. Som rettskilde nevner kandidaten Riksadvokatens rundskriv,

men har ikke den konkrete henvisningen. Kandidaten har god oversikt over bruken av hypoteser og synliggjør gjennom presentasjonen at dette er et levende materiale. Kandidaten har mange hypoteser. Nevner ulike straffebestemmelser som kan være aktuelle, strl. § 204, strl. § 201, strl. § 205, strl. § 202, og strl. § 331. Nevner hvilke av disse som er mest aktuelle i hvert fall i startfasen. Jeg synes kanskje at kandidaten presenterer litt lite teori om disse bestemmelsene og er litt for lite konkret på hvilke bevisstema i selve bestemmelsene som etterforskningen særlig må ha fokus på ...

Diskusjoner i fokusgruppene viser at det er delte meninger blant sensorene om hvorvidt man kan legge antall studiepoeng i hvert fag til grunn når karakter skal settes. Skal vurderingen til sensor i strafferett og straffeprosess veie mer enn kriminalteknikers vurdering, siden juss-faget utgjør flere studiepoeng enn det kriminaltekniske? Eller skal vurderingen alltid baseres på den helhetlige etterforskningen, uten slik vektning mellom fagene? Diskusjoner i fokusgruppene synliggjør at informantene har delte meninger om dette.

5.2.3 Kriterier og standarder brukes og oppfattes ulikt

I det følgende vil jeg redegjøre for hvordan kriterier og standarder ble brukt i vurderingene, i tråd med forståelsen som ble lagt til grunn ved nedbrytingen av dataene.

Karakterbeskrivelser:

Karakterbeskrivelsene synes å være hyppig brukt som hjelpemiddel og verktøy i karaktersettingen. Mange informanter peker på disse i sine individuelle begrunnelser, og alle fire kommisjoner støtter seg til dem i karakterdiskusjonene. Særlig begrepene *vurderingsevne* og *selvstendighet* ble gjenstand for diskusjoner i kommisjonene. I de individuelle vurderingene varierte det mer hva som ble vektlagt fra karakterbeskrivelsene. Noen brukte dem helhetlig, mens andre tok utgangspunkt i utvalgte kriterier fra karakterbeskrivelsene i sine begrunnelser, for eksempel *meget god prestasjon*, *vesentlige mangler* eller *tilfredsstillende minimumskravene*. Følgende uttalelse synes å være representativ for mange av informantene: «Disse bruker jeg mye. Jeg synes det er svært mye hjelp i disse, særlig når man totalt sett vurderer kandidatens selvstendighet og vurderingsevne».

Vurderingskriterier og læringsutbytter:

Informantenes ga ulike svar på spørsmålet om hvordan de brukte de eksplisitte vurderingskriteriene for etterforskningsemnet i sin bedømming av prestasjonen. Noen anså kriteriene som viktige og opplyste at de brukte dem aktivt, mens andre valgte å ikke bruke dem fordi de mente de var intetsigende og lite hensiktsmessige. Andre igjen blandet vurderingskriterier og karakterbeskrivelser, og svarte på spørsmålet som om det omhandlet bruken av karakterbeskrivelsene. Følgende tre sitater illustrerer spennet i informantenes oppfatninger om bruk av de eksplisitte kriteriene: «De er svært viktig for å få mest mulig lik vurdering fra alle sensorer på samme fag» og «Jeg brukte dem, men hadde nok ikke finlest dem» og «Jeg må vel innrømme at disse kommer litt i bakgrunnen. Synes de er generelle og intetsigende. Jeg får lite ut av slike kriterier». Sitatene er hentet fra informantenes svar på spørsmål om hvordan de brukte vurderingskriteriene for etterforskningsemnet.

Et mer entydig funn var at kriteriene i liten grad ble omtalt, eller vist til, i karakterbegrunnelsene. Det gjaldt både i kommisjonene og i de individuelle vurderingene. Det samme var tilfellet for læringsutbyttebeskrivelsene i emnet. Med det mener jeg at ordene læringsutbytter, læringsmål eller vurderingskriterier ikke ble konkret nevnt i informantens (friteksts)begrunnelser eller i diskusjonene i fokusgruppene. Informantenes faglige begrunnelser inneholder selvsagt momenter som er å finne igjen i læringsutbyttene og i de eksplisitte vurderingskriteriene. Funnet knyttet til at de ikke syntes å bli aktivt brukt i begrunnelser og karakterdiskusjoner.

Ikke alle begrunnet hvorfor de ikke brukte kriteriene, men de som gjorde det pekte på at de kan være hemmende for helhetsvurderingen. Det kan illustreres gjennom følgende to sitater: «Hvis man blir sittende og stirre på ett av punktene kan det få for stort fokus – både opp og ned» og «Jeg har erfaring med begge deler, og jeg har valgt å stole på magefølelsen for å si det sånn, men da med priming på kriteriene i forkant». Disse informantene støttet seg til sitt profesjonelle skjønn og implisitte kriterier for hva som kjennetegner kvalitet i etterforskning og i den konkrete (eksamens)casen. De synes å se kvalifikasjoner som noe mer enn kun formelle læringsutbytter, og vurderte prestasjonen i lys av de komplekse og integrerte kvalifikasjoner som den aktuelle etterforskningen og tilhørende arbeidsoppgaver ville krevd i det virkelige liv. Fokus var først og fremst rettet mot det store bildet, og i mindre grad mot enkeltdetaljer. Noen informanter ga også uttrykk for at de benyttet en kombinasjon mellom implisitte og eksplisitte kriterier. Det kan delvis leses ut fra det siste sitatet over, og gjennom følgende sitat: «Metodikken på å utregne karakteren er koplet både mot vurderingskriteriene

og mot opplevelsen av kandidatens prestasjoner».

Informantene som opplyste å ikke bruke vurderingskriteriene, var alle erfarne sensorer innen etterforskningsemnet med minst fire års erfaring. Blant de som opplyste at de brukte kriteriene, fant jeg ingen tydelig sammenheng mellom erfaring og bruk.

Gått med saken

Et kriterium som flere informanter benyttet seg av i vurderingen, var å stille seg spørsmål om hvordan det ville gått med saken (eksamensoppgaven) i virkeligheten. Spørsmålet ble vurdert i lys av den etterforskningen kandidaten hadde presentert for kommisjonen, og svaret ble av flere brukt som tungtveiende støtte ved karakterbedømmingen, herunder for å sette karakteren F. Følgende to sitater illustrerer bruken av kriteriet *gått med saken*:

Kriteriene er klare, men spørsmålet som er egentlig essensielt er, har han kommet i mål, har han etterforsket saken her? Det har han ikke gjort i mitt fag. Jeg har valgt å ikke gi ham bestått.

Her brukte jeg kun mitt eget hode og min egen erfaring hvor jeg spurte meg om denne kandidaten hadde kommet i mål med etterforskningen slik han presenterte den. Svaret her for meg ble nei. Det var mange etterforskningsskritt som lå helt åpne og som oppgaven innbød til at kunne lukkes.

Alle fag var representert blant informantene som benyttet tilnærmingen om hvordan det ville gått med saken. Denne typen vurdering var dominert av å være holistisk og basert på implisitte kriterier og den enkeltes profesjonelle skjønn, heller enn å være analytisk og basert på eksplisitte kriterier. Sensorene gjorde en helhetlig vurdering av hvordan saken samlet sett ville ha stilt seg dersom etterforskningen ville blitt gjennomført som den var presentert – med utgangspunkt i sine egne oppfatninger av hva som kjennetegner en god etterforskning. Vurderingene var preget av å være kompetanseorientert med fokus på de integrerte kompetanser som trengs for å utføre etterforskningsarbeid i det virkelige liv.

Kriteriet *gått med saken* er ikke direkte nedtegnet i de eksplisitte vurderingskriteriene, men kan kanskje indirekte leses ut fra dem, for eksempel gjennom formuleringen: Er oppgaven

besvart slik at etterforskningen kan være egnet for fremtidige påtalespørsmål?

Sensorveiledninger

Det hersker ulike oppfatninger blant informantene om hvordan, eller om, sensorveiledningene skal brukes i vurderingen. Følgende fire sitater illustrerer spennet i oppfatningene: «Jeg leste sensorveiledningene nøye. Brukte de under eksamen i tillegg til at de ble lest før selve eksamen» og «Jeg tok utgangspunkt i den aktuelle kommisjonssammensetningen og leste igjennom de tre mest aktuelle sensorveiledningene» og «Jeg leste sensorveiledningen for mitt eget fagelement. Jeg benyttet det ikke aktivt under presentasjonen» og «Jeg brukte ikke sensorveiledning».

Som for vurderingskriteriene og læringsutbyttene, ble også sensorveiledningene i liten grad vist til, eller omtalt, i karakterdiskusjonene i kommisjonene. De som svarte at de ikke brukte sensorveiledninger, var i hovedsak erfarne sensorer.

Pensum og undervisning:

De fleste informanter opplyser å ha lest pensum i eget fag, noen kun deler av det. Mange har også lest noe pensum i andre fag, men ingen har lest alt pensum i emnet. I informantenes karakterbegrunnelser vises tidvis til pensum, og da oftest på en overordnet måte som illustrert gjennom følgende to sitater: «Jeg vurderer kandidaten til en sterk C. Kandidaten har kontroll på en del av teorien, men evner ikke å knytte den helt til den konkrete saken» og «Kandidaten viser ikke forståelse for teorien som ble nevnt tidligere i redegjørelsen, på en praktisk god måte». I noen tilfeller ble også faglige momenter som ligger utenfor pensum vektlagt i vurderingen. Det gjaldt både faglig innhold som har vært pensum tidligere i studiet, og innhold som det har vært undervist i, men som aldri har vært pensum.

Det hersker ulike oppfatninger blant informantene om hvorvidt sakstyper som har vært gjenstand for undervisning, skal ha betydning for vurderingen. Noen mener at man skal kunne kreve mer av kandidater som skal besvare eksamensoppgaver med sakstyper som det har vært undervist i, mens andre mener at innhold i undervisningen ikke skal ha betydning for vurderingen.

5.2.4 Ulike tilnæringer i argumentasjon og overtalelse

Alle tre appellformer (etos, patos, logos) ble brukt i samspillet når karakter skulle settes – ofte i kombinasjon med hverandre. Jeg vil i det følgende gi noen eksempler på det, og peke på trekk ved argumentasjonen som syntes å ha betydning for overtalelsen – og dermed karakterfastsettelsen. Jeg vil bemerke at dataene inneholder et spekter av ulike argumenter og begrunnelser, og at ikke alle faller inn under det som omhandles her. Det som presenteres illustrerer likevel det jeg mener er tendenser, eller mønstre, i empirien.

En del uttalelser i fokusgruppene appellerte til mottakers følelser (patos) eller til forhold ved informantene (etos). Med *forhold ved informantene* mener jeg først og fremst deres erfaring, men også hvorvidt de virket sikker i sin argumentasjon og stod fast ved sitt syn. Bastante argumenter synes å ha større overtalelseskraft enn det som var tilfellet for argumenter av mer undrende eller spørrende art – særlig hvis de kom fra en erfaren sensor. Bastante argumenter var i mange tilfeller kjennetegnet av å være konkluderende, men ikke begrunnet. Følgende to sitater illustrerer det: «Selvstendighet er helt borte i min bok. Det klarer han ikke å vise» og «Kandidaten mangler fullstendig rede på hvordan ting skal sikres». I kommisjoner hvor flere karakterer skilte de individuelle vurderingene, så jeg tilfeller hvor mindre erfarne sensorer uttrykte en annen styrke på karakteren enn det som var gjort i den skriftlige, individuelle begrunnelsen. For eksempel ble en individuelt begrunnet D uttrykt som en «veldig veldig svak D» i kommisjonen, etter at erfarne kommisjonsmedlemmer hadde argumentert for at deres vurdering lå helt ned mot en F. Imidlertid var ikke karakteren D omtalt som veldig svak i den individuelle begrunnelsen.

Under vises utdrag fra diskusjonen i en fokusgruppe som illustrerer bruk av appellformene etos og patos. Fag 1 er representert ved en erfaren sensor som har vurdert prestasjonen er en A, mens fag 2 er representert ved en uerfaren sensor som har vurdert prestasjonen til en C.

Fag 1: Jeg har en del sensurerfaring og synes denne kandidaten har bedre innsikt enn mange andre. Når jeg hører på denne kandidaten så tar jeg nesten bølgen ... Jeg vil være tydelig på at jeg synes C er altfor strengt

Fag 2: Ja, nettopp, ja

Fag 1: Jeg tenker at hvis vi hadde sagt C på denne kandidaten så tenker jeg at stakkars jævel som noensinne skal få A

Fag 2: Nettopp

Her brukes argumentasjon som spiller både på talers erfaring (jeg har en del sensurerfaring) og på følelser i form av å skape begeistring (jeg tar neste bølgen) og medfølelse (stakkars jævel som noen sinne skal få A) hos vurdererne. Sensor fra fag 2 endte opp med å uttrykke usikkerhet rundt egen vurdering, og at den trolig hadde vært for streng. Selvsagt trenger ikke det å være negativt. Diskusjoner med kolleger kan gjøre at man ser ting i et nytt lys, og det er positivt at man ikke er låst i eget syn. Jeg vil understreke at mitt formål ikke er å mene noe om karakterbedømmingen, men å se på trekk ved (i dette tilfellet) kommunikasjonen og argumentasjonen som synes å kunne ha betydning i samspillet når karakter skal settes.

Et annet eksempel som illustrerer bruk av patos, er dette:

Når kandidaten bommer såpass kraftig på tvangsmiddelbruken så er jeg litt redd for at vedkommende vil bomme kraftig på en del andre essensielle ting. Og ville kandidaten klart å bruke den jussen inn i et avhør for eksempel? Det stiller jeg store spørsmål ved.

Her gjøres det antagelser om kandidatens kompetanser ved å benytte patos (redd for) i kombinasjon med etos (jeg).

Betydningen av den innledende kommunikasjonen i kommisjonene

Kommisjonsmedlemmene i fokusgruppene hadde ulike tilnærminger til hvordan de innledet karakterdiskusjonene. Noen innledet med å snakke om prestasjonens styrker og svakheter uten å nevne karakter, mens andre begynte med å fortelle hvilken karakter den enkelte hadde vurdert prestasjonen til. De ulike tilnærmingene syntes å få betydning for den videre diskusjonen i kommisjonen. De fokusgrupper som innledet med å snakke «ned» prestasjonen, og appellere til følelsen (patos) av at den var dårlig, ble også i fortsettelsen preget av et større fokus på det som var negativt, enn for de kommisjoner hvor man innledningsvis også påpekte styrker ved prestasjonen. Under vises eksempler på ulike innledninger ved vurdering av samme prestasjon. Den videre karakterdiskusjonen i kommisjon 1 hadde jevnt over større fokus på svakheter ved prestasjonen, enn det som var tilfellet for kommisjon 2.

Innledning kommisjon 1:

Fag 1: Jeg synes det er veldig ryddig å begynne med bare å si hva utgangspunktet var for vurderingen vår, slik at vi ser hva avstanden er før vi begynner å.. Hvem har lyst til å begynne?

Fag 2: Jeg kan begynne. Jeg har strøket denne kandidaten.

Fag 1: Jeg har latt kandidaten stå på en E og den er syltynn.

Fag 3: Jeg var på en..jeg gikk til D fordi..ja, jeg kan ta det etterpå. Men en veldig dårlig D.

Innledning kommisjon 2:

Fag 1: For mitt fag så kan jeg starte med å ta litt innledningsvis, og det som var bra for meg var å høre i den besvarelsen her som kandidaten redegjorde for, det var hans hypotesetinking hvor han grupperte inn i fire ulike bokser med blant annet den brannårsaken som var påsatt, frivillig påsatt, uaktsom, teknisk feil og naturlig årsak og gjorde rede for de hypotesene. Jeg forventet da han skulle komme inn på de hypotesene senere og etterforske de, og det er vel der at jeg skal diskutere litt med dere ift den videre gangen ...

Uttalelsene til informanten som innledet diskusjonen, syntes også å legge føringer for hva kommisjonens øvrige medlemmer fokuserte på i sine første kommentarer. I kommisjon 1 fulgte de to andre sensorene opp med å kortfattet uttrykke hvilken karakter de hadde landet på – etter samme «mal» som førstemann. I kommisjon 2, derimot, fulgte sensorene opp med å redegjøre mer utførlig for sine vurderinger av prestasjonens styrker og svakheter, i tråd med det førstemann hadde gjort. Karakterdiskusjonens innledning kan dermed ha betydning for innholdet i samtalen mellom sensorene.

Generelt fant jeg at det var større fokus på det positive ved prestasjonen hos en faglig sterk kandidat enn for en svakere kandidat. Informantene brukte også kortere tid på å vurdere den sterke kandidaten⁷. Kommisjonene som vurderte den sterkeste kandidaten brukte henholdsvis 12 og 16 minutter på å enes om karakter, mens kommisjonene som vurderte den svakere

⁷ Den kandidaten som fikk best karakter av de to. Den sterkeste ble vurdert i spennet mellom A og C/D, den svakeste mellom C/D og F.

kandidaten brukte 26 og 27 minutter.

Generelle begrunnelser som grunnlag for karakterfastsetting

Logos handler om argumentasjonens oppbygging. En utfordring ved en del av de argumentene og slutningene som ble fremsatt i kommisjonene, og som fikk betydning for den endelige karakteren, var at de var generelle og lite utdypet. Det samme var tilfellet for de individuelle begrunnelsene. Med det mener jeg at de manglet informasjon om *hva* ved prestasjonen som understøttet dem. Dermed var det tidvis vanskelig å forstå hva de konkret innebar, herunder hva ved prestasjonen som var bra eller mindre bra, og hvorfor. Under vises noen uttalelser som illustrerer dette:

«Generelt gikk kandidaten i liten grad inn og knyttet vurderingstemaene i loven opp mot de nødvendige straffeprosessuelle skrittene og det etterforskningstaktiske, det var kanskje det som skuffet meg aller mest».

«Altså mitt fag isolert sett klarer han seg til en viss grad på, med svakheter, men han klarer jo ikke å knytte det sammen med resten og heller ikke... alle etterforskningstingene kommer jo egentlig ikke og blir jo ikke knyttet inn».

«Kandidaten treffer ikke på selvstendighet og får i liten grad frem forståelse selv når han ledes til å forklare skjønnstemaene for de aktuelle bestemmelser».

«Jeg synes ikke kandidaten klarte å visualisere noen form for målstyrt etterforskning».

«Kandidaten ble for generell i oppramsingen av de forskjellige etterforskningsverktøyene slik at den informasjonen han satt med utfra oppgaveteksten, ikke ble målrettet nok».

Informantene syntes å være fortrolige med slike generelle formuleringer, og etterspurte sjelden utdypning av argumentene, eller mer presist slutningene, som de må kunne sies å ha karakter av. I mange tilfeller sluttet informantene seg til dem og la dem til grunn for karakterfastsettelsen. Det er vanskelig å si noe sikkert om hvordan de ble vektet i vurderingen, men det er ingen tvil om at de ble vektlagt. Forhold som sikkerhet og grad av bastantheit i argumentasjonen syntes, som tidligere omhandlet, å ha betydning ved vektingen. En

utfordring ved denne typen argumentasjon er at den er lite tilgjengelig og vanskelig å etterprøve – selv om informantene syntes å forstå, eller i alle fall akseptere, argumentene. Hva innebærer det egentlig at kandidaten ikke treffer på selvstendighet, eller at vedkommende ikke klarer å visualisere en målstyrt etterforskning?

5.3 Refleksjoner rundt funnenes generaliserbarhet

Generaliserbarhet er et sentralt kvalitetskriterium innen kvalitativ forskning, og handler om forskningens relevans utover de involverte forskningsdeltakere i den aktuelle konteksten. (Tjora, 2017, s. 231). Det er viktig å reflektere over funnenes mulige overførbarhet, fordi de i så fall vil kunne være retningsgivende for andre kontekster og objekter, og danne grunnlag for kvalifiserte beslutninger også ved utdanningsinstitusjoner utover PHS. Gjennom rapporteringen har jeg tilstrebet å være transparent og detaljert i mine beskrivelser av forskningen, og håper med det å ha lagt grunnlag for at leseren selv kan reflektere rundt spørsmålet om generaliserbarhet. Tjora kaller dette for naturalistisk generalisering (Tjora, 2017, s. 238).

I prosjektet har jeg samlet data fra 29 informanter. De har gitt utfyllende og detaljerte begrunnelser for sine vurderinger av en muntlig eksamensprestasjon gjennom å besvare spørsmål i et spørreskjema og gjennom fokusgruppeintervju. Som jeg tidligere har beskrevet, er kunnskap kontekstuell, og vil ikke automatisk kunne overføres til, eller sammenlignes med, en annen situasjon. Et mål med prosjektet har vært å se på mer generelle trekk og karakteristika ved vurderingene, heller enn det rent faglige. I analysen kodet jeg dataene analytisk heller enn deskriptivt, fordi en slik tilnærming, ifølge Gibbs, kan være til hjelp for å forstå dataene på måter som strekker seg utover informantenes uttalelser og som blir mer generelle (2018, s. 54, 73). Mine funn omhandler mer overordnede forhold ved kommunikasjonen, for eksempel at begrunnelsene tidvis er generelle, at fokus er ulikt og at innledningen til karakterdiskusjonen synes å ha betydning for den videre samtalen. Funnene omhandler også, mer overordnet, hvordan vurderingskriterier, pensum og karakterbeskrivelser brukes på tvers av sensorer. Slik jeg ser det, er dette forhold som kan belyse noe mer enn det som er utelukkende subjekt- og kontekstbasert, og som vil kunne være interessant i vurderingsarbeid generelt. Disse funnene vil i så fall kunne brukes også i andre emner ved PHS og av fagpersonell ved andre utdanningsinstitusjoner, i arbeidet med å bedre tilrettelegge

for rettferdig og riktig vurdering av studentprestasjoner. For eksempel kan kunnskapen bidra til økt fokus rundt karakterdiskusjonenes innledning.

6. Drøfting

Et hovedfunn i undersøkelsen er at sensorene vurderer ulikt, og at de verdsetter og vektlegger ulike forhold ved samme prestasjon. Funnet er i seg selv ikke veldig oppsiktsvekkende, og har bred støtte i forskning (se for eksempel Asmyhr, 2011; Harlen, 2005; Raaheim, 2000; Rye, 2014). I det femdimensjonale rammeverket for autentisk vurdering fremheves subjektivitet som et grunnleggende element som gjennomsyrer alle dimensjonene, herunder oppfatninger av hva som er autentisk og hvordan ulike kriterier og standarder skal forstås (Gulikers et al., 2004, s. 69). Siden all vurdering innebærer elementer av skjønn, er det vanskelig, om ikke umulig, å si sikkert hva som er «riktig» karakter for en eksamensbesvarelse (Raaheim, 2016, s. 41). Eksamen i etterforskningsemnet er kompleks og omfattende, og det er mulig å vektlegge mange ulike forhold ved det presterte. Hva den enkelte vektlegger og verdsetter vil, ifølge Eisner, henge sammen med hva vedkommende selv kan si noe om, som igjen henger sammen med personens erfaringer og tradisjoner (2017, s. 7-8, 46).

I et studie av muntlige eksamensprestasjoner ved et norsk masterstudium, fant Dobson at sensorenes bakgrunnsinteresser og kompetanser – sammen med tema for oppgaven – la føringer for hva som ble eksamenssituasjonens dominerende hovednarrativ (2007, s. 142). Det er forenlig med mine funn om at vurderingene i stor grad tar utgangspunkt i sensors fagtilknytning. I PHS' rapport fra sensurprosjektet ble vurderernes ulike fag- og yrkesbakgrunner nettopp trukket frem som en potensiell utfordring, fordi det kan gi seg utslag i (for) stor grad av subjektivitet i vurderingsarbeidet (Politihøgskolen, 2016b, s. 92). Friche (2011) har analysert vurderingspraksis blant lærerteam innen fire ulike praksisvirksomheter (smed, klinikk, gastronom og mediegrafikk), og fant at det hersker ulike (konstruerte) logikker for vurdering av elevprestasjoner på tvers av yrkesutdanninger. Mens noen ser det som viktig å vektlegge egenskaper ved et konkret studentprodukt, vektlegger andre studentenes evne til refleksjon over styrker og svakheter ved produktet. De ulike logikkene skyldes forskjellige oppfatninger av hva som kjennetegner utdanningskvalitet, og påvirkes av blant annet yrkets utdanningskultur og historikk. Sensorer med ulik fagbakgrunn kan dermed ha vidt forskjellige oppfatninger av hva som kjennetegner kvalitet i et studentprodukt (Friche,

2011). I etterforskningsemnet representerer ikke sensorene bare ulike fag, men også ulike profesjoner, for eksempel politi og juss. De har ulike bakgrunner, erfaringer og faglige kompetanser, som vil prege deres subjektive oppfatninger om hva som bør tillegges vekt.

I en undersøkelse ved grunnfag psykologi ved Universitetet i Bergen, fant Raaheim at det var store avvik i vurderinger på tvers av sensorer (Raaheim, 2000). I Raaheims studie var det ikke utarbeidet prosedyrer for rettleiing, for eksempel i form av sensorveiledninger. Derfor, skriver han, er det kanskje ikke så rart at vurderingene ble ulike. En av de mest erfarne sensorene uttrykte at «dette åpner opp for et skjønn hos vurdererne hinsides all rimelighet» (Raaheim, 2000, s. 209). I mitt studie, derimot, er det utarbeidet eksplisitte vurderingskriterier og sensorveiledninger. I tillegg setter pensum og læringsutbytter rammer for hva som skal vurderes og vektlegges. Likevel oppfattes og brukes disse ulikt, og blir i liten grad vist til i karakterbegrunnelsene. Det er altså ingen automatisk sammenheng mellom utarbeidelsen av kriterier, og om, eller hvordan, de brukes. I det femdimensjonale rammeverket fremheves kriterier og standarder som viktige for autentisiteten, under forutsetning av at de er relevante og virkelighetsnære, og at de gjøres eksplisitte og transparente i forkant av vurderingssituasjonen (Gulikers et al., 2004, s. 75). Også Harlen og Gynnild peker på at en felles forståelse av kriterier og læringsmål er nødvendig for å øke reliabiliteten i vurderingene (Gynnild, 2010, s. 7; Harlen, 2005, s. 267). Mine funn kan tyde på at det ikke eksisterer slik felles forståelse blant sensorene. Dermed åpnes det kanskje også i større grad for ulik bruk av kriteriene. Det er forenlig med et studie gjennomført av forskerne bak det femdimensjonale rammeverket, hvor et formål var å teste om rammeverket var komplett eller hadde mangler. De fant at det ikke tar høyde for vurderernes sentrale betydning i autentisk vurdering, og at sensorene må involveres i arbeidet med å definere og ta i bruk kriterier og standarder (Gulikers et al., 2004, s. 76-83). En hensiktsmessig utvidelse av rammeverket vil derfor være å innføre en sjette dimensjon, *vurdereren*, eventuelt å innta denne som et underpunkt til dimensjonen om standarder og kriterier (Gulikers et al., 2004, s. 82).

En annen mulig årsak til den ulike bruken av kriterier, kan knyttes til kompleksiteten i etterforskningsemnet. Eisners kritikk mot standardisert testing rettes blant annet mot at slike tester opererer med forenklete og avgrensede valg og tolkningsmuligheter, og at de dermed er «lette» å vurdere (Eisner, 2017, s. 47). Snur man dette på hodet kan man si at etterforskningseksamen representerer det stikk motsatte, hvor valgmulighetene over hva som kan vektlegges er mange. Bare skrivet med faglige vurderingskriterier er 13 sider langt. Ved

vurdering av komplekse oppgaveutførelser vil det alltid være utfordringer forbundet med at kriterier og standarder ikke er direkte målbare, og at det profesjonelle skjønn blir et tolkende mellomledd i vurderingen (Gynnild, 2010, s. 13). Gynnild forklarer ulik bruk av kriterier i lys av Dreyfus og Dreyfus' modell for ferdighetstilegnelse (fra novise til ekspert). For en nybegynner er regler og retningslinjer helt sentrale for å kunne utføre arbeidet. Etter hvert vil regelstyringen avta til fordel for intuitive beslutninger som bygger på erfaring (Dreyfus & Dreyfus, 2012; Gynnild, 2010, s. 10). Det er forenlig med mine funn, som viser at alle informantene som opplyste at de ikke støttet seg til de eksplisitte vurderingskriteriene, var erfarne med minimum fire års erfaring som sensor i emnet. Det betyr imidlertid ikke at faglige kriterier, eller læringsutbyttene, ikke ble brukt. Trolig er det heller et uttrykk for at sensorene dro veksler på disse som en implisitt forståelse av faglige krav gjennom utøvelse av sitt profesjonelle skjønn. Kriterier og standarder er nemlig helt nødvendige betingelser i all prestasjonsvurdering, enten de benyttes eksplisitt eller implisitt (Gynnild, 2013, s. 37). Som jeg vil komme tilbake til, er den ulike bruken likevel en utfordring for reliabiliteten og validiteten i vurderingene.

Ved Norges musikkhøgskole (NMH), erfarte man, i et pilotprosjekt, å oppnå mer konsistente vurderinger av sangprestasjoner, etter utvikling av eksplisitte kriterier for analytisk vurdering (Gynnild, 2010). Bakgrunnen for prosjektet var erkjennelsen av at kunstneriske prestasjoner er komplekse og nyanserte, og vil kunne bedømmes svært ulikt avhengig av øyet som ser (Gynnild, 2010, s. 15). I etterforskningsemnet legges det opp til en kombinasjon mellom holistisk og analytisk vurdering, slik jeg ser det. På den ene siden bes kandidatene om å besvare, eller redegjøre for, en helhetlig etterforskning, noe som tilsier at prestasjonen skal vurderes helhetlig. På den andre siden er det utarbeidet kriterier i emnet, med utgangspunkt i læringsutbyttene, som beskriver enkeltdeler ved etterforskning som skal, eller kan, vurderes. I tillegg er emnet faginndelt, med sensorer som representerer ulike fag i kommisjonene. Mine funn viser at mange sensorer støtter seg til sitt profesjonelle skjønn i vurderingen, og at bruken av de eksplisitte kriteriene varierer. Dobson betegner kombinasjonen mellom analytisk og holistisk tilnærming som «en bevegelse bakover og fremover mellom ytre narrativer og «in-the-head» normative synsvinkler» (2007, s. 144). En erfaring fra prosjektet ved NMH var at en for «nærsynt» bruk av kriterier neppe bidrar til mer pålitelig eller gyldig vurdering. Utfordringen lå i å finne den krevende balansen mellom deler og helhet (Gynnild, 2010, s. 11). Det er i tråd med uttalelser fra flere av mine informanter, som opplevde kriteriene som hemmende for helhetsvurderingen. Det kan godt tenkes at en prestasjon som samlet sett

fremstår fremragende, ikke nødvendigvis får full score på enkeltdelene, eller at en prestasjon som får høy score på enkeltdelene, ikke nødvendigvis fremstår fremragende samlet sett (Sadler, 2009b, s. 165). Tross gode erfaringer ved NMH, stiller Gynnild spørsmål ved om analytisk vurdering av komplekse utfordringer overhodet er mulig, fordi komplekse kompetanser aldri vil kunne settes på formel. En analytisk tilnærming vil alltid være reduksjonistisk fordi det ikke er mulig å utvikle kriterier som fanger alle nyanser (Gynnild, 2010, s. 15). Heller ikke Sadler mener det er mulig å redusere kvalitative vurderinger til et sett prosedyrer, eller kriterier, som i neste omgang skal brukes og forstås likt av studenter og sensorer (Sadler, 2009a, s. 820). Vurderingen gjøres av mennesker i konkrete situasjoner, og subjektivitet vil alltid ha betydning (Eisner, 2017).

I min undersøkelse var karakterbeskrivelsene desidert mest brukt av de tilgjengelige kriterier og standarder. De fleste opplevde å ha nytte av slike mer overordnede, eller vide, kriterier. Noen støttet seg også til andre typer overordnede kriterier, herunder å stille seg spørsmål om hvordan det ville gått med saken i virkeligheten. Disse kriteriene og standardene er betydelig mindre komplekse og oppdelt, og trolig mer håndterlige, enn det 13 siders lange skrivet med eksplisitte kriterier, eller de fem sensorveiledningene som er utarbeidet for hver oppgave. I lys av Mulders (2014) oppfatninger av kompetanse, synes *gått med saken*-tilnærmingen å være kompetanseorientert heller enn kompetansebasert. Det er fordi en slik tilnærming retter større fokus mot kvalifikasjonenes relevans for arbeidslivet og ikke (bare) på de formelle kvalifikasjonene slik de er beskrevet gjennom læringsutbyttene. Innenfor en slik tilnærming vil en kvalifikasjon kunne ses som noe mer enn et formelt læringsutbytte, som er Kunnskapsdepartementets forståelse av begrepet (2014, s. 35). Tilnærmingen er mer i samsvar med en holistisk tilnærming og det Faultley (2010) kaller top-down-strategi, hvor kompetanse forstås som et integrert konsept og hvor mange kriterier kan ligge til grunn for helhetsvurderingen av prestasjonen. Vurdererne opptrer mer som kvalifiserte etterforskere i bedømmingen, og vurderer prestasjonen ut fra eget profesjonelle skjønn og oppfatninger av hva som kjennetegner kvalitet. En slik tilnærming vanskeliggjør bruk av eksplisitte kriterier. Funnet om at en del begrunnelser var generelle og lite utdypet, kan kanskje forstås i lys av dette. I en holistisk vurdering vil det først og fremst være de store linjene som danner grunnlag for karakterbedømmingen. Dermed er det ikke sikkert at prestasjonen er vurdert i et detaljnivå som slike konkrete begrunnelser krever. Sadler skriver at det er et veletablert faktum at eksperter kan gjenkjenne kvalitet når de ser det, også i tilfeller hvor de ikke kan forklare det med ord (Sadler, 2009a, s. 820).

Sensorenes hyppige bruk av kriteriene i karakterbeskrivelsene, er i tråd med UHRs operasjonalisering av kriteriebasert vurdering, nemlig at «dersom en prestasjon tilfredsstillende kriteriene for en karakter skal man gi denne karakteren uavhengig av hvordan fordelingen av de øvrige karakterene i eksamenskullet er» (Universitets- og høyskolerådet, 2008, s. 3). Imidlertid mener Gynnild at UHRs formulering er problematisk, fordi den til dels springer ut av et normbasert grunnlag som står i motstrid til målsetningen om at karakteren skal speile kandidatens kompetanse. Det å ta utgangspunkt i *kriteriene for en karakter* er noe annet enn å ta utgangspunkt i faglige kriterier og standarder – som er en målsetning i Nasjonalt kvalifikasjonsrammeverk (Kunnskapsdepartementet, 2014). Vurderingen må alltid springe ut av faglige kriterier og standarder, og så må karakteren i neste omgang speile denne kompetansen på en riktig måte (Gynnild, 2013, s. 39). Den utstrakte bruken av karakterbeskrivelsene kan dermed være til hinder for en vurdering som tar utgangspunkt i kandidatens oppnådde kompetanser. All den tid UHR ikke har utdypet eller forklart hvordan prinsippene for kriteriebasert vurdering skal omsettes i praksis, er det opp til sensor å selv definere hva kriteriebasert vurdering betyr (Gynnild, 2013, s. 40).

Siden autentisitet er subjektivt og kan endres med erfaring, er det naturlig å stille større krav til autentisitet i et avgangskull enn for førsteårsstudenter (Gulikers et al., 2004, s. 76). Når studentene møter til eksamen i etterforskning er de i slutten av sitt siste studieår, og kan få uker senere være ansatt i et politidistrikt og utføre etterforskningsoppgaver i arbeidslivet. Det taler for at eksamen i etterforskning bør tilstrebe høy grad av autentisitet. Ifølge rammeverket, og i tråd med målsetningene i Nasjonalt kvalifikasjonsrammeverk, bør en autentisk oppgave integrere testing av de kompetanser som kreves for å utføre tilsvarende oppgaver i virkeligheten. Den fysiske og sosiale konteksten bør være gjenkjennbar sammenlignet med profesjonell praksis, og vurderingsresultatet (besvarelsen) må være av en slik art at studentenes kompetanser kan bedømmes (Gulikers et al., 2004, s. 71-75). I lys av ovennevnte ser jeg flere utfordringer knyttet til dimensjonene i rammeverket. Til eksamen i etterforskningsemnet bes kandidatene om å redegjøre for en helhetlig og formålsstyrt etterforskning av et mulig straffbart forhold som de får beskrevet. Det er et mål at redegjørelsen (etterforskningen) skal være i tråd med gjeldende metodikk, regelverk og kunnskapsgrunnlag. Oppgaven legger opp til relevant og virkelighetsnær tankevirksomhet om blant annet hva som kan ha skjedd og hva som bør gjøres videre. Imidlertid legger verken oppgaven eller konteksten opp til testing av sentrale ferdighetsmål i emnet, for eksempel utarbeidelse av etterforskningsplan i saken, eller planlegging, gjennomføring og

dokumentering av etterforskningsmetoder. Kandidaten får ingen saksdokumenter eller filer med digitale spor, som man ville hatt tilgang til i en reell etterforskning. De får ikke tilgang til relevante digitale verktøy og ressurser, og de har ikke anledning til å samarbeide med de aktører som oppgaven legger opp til – som påtalejurist og fagspesialister. Studentene har en time til å forberede seg, og deretter tre kvarter til å presentere sin etterforskning. I en reell etterforskning vil man ha tid til å tenke igjennom, planlegge og diskutere med andre. I løpet av timen studentene har til forberedelser, skal de ikke bare gjøre seg kjent med informasjonsgrunnlaget og lage en plan for hvordan de vil presentere etterforskningen, de skal også begrunne sine valg i lys av teori fra fem ulike fag som skal presenteres i en helhet.

Mine funn, sammenholdt med eksamensformen, gir grunn til å stille spørsmål ved reliabiliteten og validiteten i vurderingene. Hvor pålitelige er de, og måler vi det vi tror vi måler? Reliabilitet og validitet henger sammen, og i den grad vurderinger er upålitelige, vil også gyldigheten svekkes (Sadler, 2009b, s. 162). De store individuelle forskjellene i (karakter)vurderingen i etterforskningsemnet, også innad i samme fag, er en utfordring for reliabiliteten. Det samme er ulik bruk av pensum, sensorveiledninger og kriterier. Imidlertid kan det tenkes at forskjellene utjevnes ved at man benytter flere sensorer i kommisjonene. Det kan støttes av mine funn, som viser betydelig mindre variasjon i kommisjonskarakterene enn det som var tilfellet for de individuelle vurderingene. Imidlertid mener Raaheim at flere sensorer i seg selv ikke er en garanti for likhet, selv om hensikten med å bruke flere sensorer nettopp er å hindre for store utslag av subjektivt skjønn og å sikre større likhet (reliabilitet) i vurderingene (Raaheim, 2016, s. 41). All den tid de individuelle vurderingene spriker såpass mye, vil karakteren kunne variere betydelig også på tvers av kommisjoner. Det er fordi kommisjonskarakteren synes å bli satt midt mellom de individuelle vurderingene. Dermed vil den enkeltes bedømming kunne få stor betydning for total karakteren. Tre sensorer som møtes med utgangspunkt D, D og E vil komme frem til et helt annet resultat enn for utgangspunktet E, F og F. Det er forenlig med funn gjort av Raaheim (2000) i en studie om interbedømmer reliabilitet. Der fant han en klar kommisjonseffekt på karakteren som følge av at karakteren ble satt som et snitt av sensorenes individuelle vurderinger. Eksempelvis ville en kandidat, som av opprinnelig kommisjon fikk karakter 2,7, oppnådd karakter 1,9 om besvarelsen hadde blitt bedømt av en annen kommisjon.

Et sentralt spørsmål knyttet til validitet er om det rent faktisk er mulig, ut fra denne ene store eksamenen, å fatte gyldige slutninger omkring studentenes etterforskningskompetanser. Ofte

vil det ikke være mulig å fatte gyldige og rettferdige slutninger om en students kompetanser på bakgrunn av én prøve. Kompetansebedømminger må baseres på flere ulike vurderinger, i ulike kontekster og ved bruk av ulike vurderingsformer (Darling-Hammond & Snyder, 2000, s. 527; Gulikers et al., 2004, s. 73, 75). De nevnte kontekstuelle ulikhetene mellom etterforskningseksamen og en reell etterforskning, kan utfordre innholdsvaliditeten. Er eksamensoppgaven virkelighetsnær nok for å måle læringsutbyttene, og får studentene anledning til å demonstrere sine etterforskningskompetanser gjennom denne eksamensformen? En annen validitetsproblematikk knyttes til funnet om at de individuelle vurderingene i stor grad ble gjort med utgangspunkt i faget som sensor representerer. I fokusgruppene ble denne tilnærmingen begrunnet med at helheten uansett ville bli ivaretatt i kommisjonene. Helhetsvurderingen i kommisjonene var også den viktigste årsaken til at informantene var villige til å fravike sin opprinnelige karakter. Det kan imidlertid vise seg å være en feilslutning, når man ser funnene samlet sett. De viser at de individuelle vurderingene i stor grad gjøres med fagspesifikt utgangspunkt og at kommisjonskarakteren settes som et snitt av karaktervurderingene for tre av de fem fagene som inngår i emnet. Kun et fåtall sensorer har lest pensum utover eget fag, og ingen har lest alt pensum. I tillegg er det svært varierende hvorvidt sensorene støtter seg til sensorveiledning(er) ved vurderingen. På bakgrunn av ovennevnte synes det å være et misforhold mellom det studentene blir bedt om å besvare (en helhetlig etterforskning), og det de blir vurdert i. Det gir grunn til å stille spørsmål ved om vi måler det vi ønsker å måle (konstruktvaliditet), altså om karaktersettingen gjenspeiler studentenes helhetlige etterforskningskompetanse. Dersom karakteren gir et uriktig bilde av studentens kompetanse, utfordres konsekvensvaliditeten, fordi karakteren kan få betydning for videre muligheter i studier og arbeidsliv. Konsekvensvaliditeten kan også påvirkes dersom vurderingen fokuserer på noe annet enn det studentene har brukt tid på i læringsarbeidet. I etterforskningsemnet har oppgaveteksten til eksamen (redegjør for en helhetlig og formålsstyrt etterforskning) vært kjent for studentene gjennom hele undervisningsperioden, og det er nærliggende å tro at den har hatt betydning for hva de har brukt (og ikke brukt) tid på frem mot eksamen. Det store spennet i karakterbedømminger av samme prestasjon tyder på at karaktersettingen kan være urettferdig. Dermed utfordres også strukturvaliditeten, fordi karakterbedømmingen synes å avhenge av hvilke sensorer og hvilke fag som er representert i kommisjonen. Tilsvarende vil den utstrakte bruken av implisitte kriterier kunne utfordre den strukturelle validiteten, fordi bruken er skjult og personlig heller enn åpen og transparent (Baartman et al., 2007, s. 123-125).

Et autentisk vurderingsresultat (eksamensprestasjon) bør innrettes slik at man kan fatte gyldige slutninger om kandidatens underliggende kompetanser (Gulikers et al., 2004, s. 75). Eisner fremhever *sammenheng* i argumentasjonen, og mellom argumenter og slutning, som en viktig faktor for å kunne vurdere gyldighet og troverdighet av et kvalitativt produkt (Eisner, 2017, s. 53). Disse «idealene» står i delvis motstrid til retoriske ytringer, hvor sentrale argumenter ofte forblir usagt. Logos viser til argumentasjonen i talen, og dennes reelle eller tilsynelatende beviskraft (Eide, 2006, s. 28). Taleren mobiliserer logos når han gjennom overbevisende argumenter, herunder de uuttalte og implisitte, viser hva som er sant eller sannsynlig (Andersen, 1995, s. 34; Johannessen et al., 2018, s. 201). Mine funn viser at flere av de argumenter og delslutninger som ble lagt til grunn for karakteren, var generelle og lite konkretiserte. Det betyr ikke at det ikke var en logisk sammenheng mellom dem, eller at de ikke var egnet til å overbevise. For eksempel var det logisk sammenheng mellom en svak karakter og argumentet om at bevisemaene var mangelfulle, eller at etterforskningen var lite målstyrt. Utfordringen lå i den uuttalte informasjonen om *hva* som gjorde bevisemaene svake eller etterforskningen lite målstyrt. Det kan utfordre validiteten, fordi en valid tolkning fordrer at den støttes av tilstrekkelige bevis (Baartman et al., 2007, s. 119; Kane, 2016, s. 198). Det utfordrer også reliabiliteten, fordi det ikke er mulig å fullt ut forstå de konkrete premisser som ligger til grunn for karakterbedømmingen. Funnet om lite utdypede argumenter, er forenlig med funn gjort av Rye (2014). Han undersøkte karaktersetting og tilhørende begrunnelser på bacheloroppgaver i sosiologifaget, hvor seks sensorer – hver for seg og kollektivt – vurderte tolv bacheloroppgaver fra seks læresteder. Han fant at karaktersettingen var preget av tilfeldigheter, både i karakterer, begrunnelser og i bruken av kriterier (s. 75). Blant annet ble det brukt argumenter som «ikke var klart objektivt definerte, og som vanskelig lar seg forklare eksplisitt og dermed også blir vanskelig tilgjengelig og etterprøvbare» (s. 72). Oppgavene kunne for eksempel bli omtalt som *spenstige* eller tekstene som *modne* eller *umodne*, uten at det ble utdypet hvorfor man mente dette. I tråd med mine funn, fant også Rye at sensorene syntes å være fortrolige med slik generell argumentasjon.

Når slutninger baseres på premisser som ikke er utdypet, utfordres reliabiliteten fordi etterprøving vanskeligjøres. Det vil også være vanskelig å benytte Eisners metode for *strukturell underbygging* for å kunne si seg enig eller uenig i argumentasjonen som fremmes. Strukturell underbygging fordrer argumenter som peker på og sammenstiller de biter av informasjon som underbygger slutningen (Eisner, 2005, s. 46). Selv om argumentasjonen i kommisjonene tidvis var generell og manglet strukturell underbygging, var den like fullt egnet

til å overtale og å skape konsensus. Sensorene etterspurte sjelden utdypning av argumentene, men sluttet seg ofte til dem og la dem til grunn for karakterfastsettelsen. Ifølge Eisner må det forventes at grad av presisjon i oppnådd konsensus reduseres jo mer komplekst det som skal vurderes er. Hvilket nivå man skal legge seg på for å kreve konsensus, vil avhenge av hva som bedømmes og omfanget av de konsekvenser beslutningen får (Eisner, 2017, s. 57-58). Etterforskningsemnet er komplekst og omfattende, og det vil være nærmest umulig å gå i detalj i alle begrunnelser. Fra kandidaten forlater rommet og til neste kandidat står på døren, har sensorene rundt 10-15 minutter til å bedømme prestasjonen. Vurderingene må gjøres ut fra det de sitter igjen med fra de forutgående 45 minuttene, og de notater de eventuelt har gjort underveis. Disse kontekstuelle faktorene «tvinger» sensorene til å komme frem til konsensus på kort tid. Dermed bestemmes nivået for konsensus i stor grad av kontekstuelle forhold. En digresjon i denne sammenheng er at min testperson – som er en erfaren fagperson og sensor – karaktersatte eksamensprestasjonen strengere enn det noen av mine informanter gjorde.⁸ I motsetning til informantene, hadde testpersonen mulighet til å pause og spole tilbake i videoen, noe han benyttet gjentatte ganger. I tillegg tok han utfyllende notater fra presentasjonen. Kan det tenkes at en mer inngående gjennomgang av prestasjonene ville ført til andre vurderinger enn dem jeg har funnet? Etterprøving vil uansett være en utfordring, all den tid eksamen ikke dokumenteres med lyd eller video.

En forklaring på at sensorene i liten grad krevde utdypning av hverandres argumenter, kan være at de var enige eller ikke hadde sterke motforestillinger. Mennesker som har liten eller ingen motstand mot et tema lar seg lettere overbevise enn de med sterk motstand (Eisner, 2005, s. 57). En annen forklaring kan være at sensorene stoler på sine kollegers vurderinger og legger dem (ukritisk) til grunn som korrekte. I lys av funnet om at sensorene som hovedregel tar utgangspunkt i eget fag, vil vedkommende kunne sies å være den mest kvalifiserte av de tre til å uttale seg fra sitt perspektiv. Det vil, i seg selv, kunne gi vedkommende etos og dermed overbevisningskraft i argumentasjonen. Talerens troverdighet, eller etos, har stor betydning for talens overbevisningskraft, og påvirkes både av det inntrykket publikum har av taleren fra før, og av den troverdigheten vedkommende skaper gjennom sin argumentasjon og fremtoning (Johannessen et al., 2018, s. 193-195). Det er forenlig med mine funn om at bastante argumenter og slutninger synes å ha stor overtalelseskraft – særlig hvis de ble uttalt av en erfaren sensor.

⁸ Testpersonen vurderte eksamensbesvarelse 1 til karakter D (øvrige vurderinger var i spennet A til C/D)

I min undersøkelse fant jeg at de kommisjoner som innledet diskusjonen med å peke på svakheter ved prestasjonen, generelt hadde større fokus på det som trakk ned, enn de kommisjoner som innledet med å (også) peke på positive forhold ved prestasjonen. En mulig forklaring kan være at innledningen fremkalte bestemte følelser hos sensorene – positive eller negative – som igjen fikk betydning for den videre dialogen. Patos handler om å appellere til publikums følelser som et ledd i overtalelsen, fordi våre oppfatninger og beslutninger påvirkes av vår sinnsstemning (Eide, 2006; Johannessen et al., 2018, s. 197). Eksempelvis kan diskusjonen som ble innledet med at to sensorer uttrykte at kandidaten hadde prestert til henholdsvis stryk og en syltynn E, ha utløst negative følelser i gruppen som fikk betydning for videre dialog. Sensoren som (individuellt) hadde vurdert prestasjonen til en D, fulgte opp med å uttrykke at det var en «veldig dårlig D» – uten at det var uttrykt i den individuelle begrunnelsen.

Generelt fant jeg at det syntes å være større fokus på det som trakk karakteren opp hos en faglig sterk kandidat, sammenlignet med en faglig svak(ere) kandidat. Det er kanskje helt naturlig, fordi premissene som underbygger en sterk karakter er andre enn dem som underbygger en svak karakter. Imidlertid skal en autentisk eksamensprestasjon være innrettet slik at sensorene kan fatte gyldige slutninger om kandidatens underliggende kompetanser (Gulikers et al., 2004, s. 75). Det fordrer at sensorene vurderer både styrker og svakheter ved prestasjonen – på en balansert måte. En svak kandidat har like mye krav på å bli premiert for sine styrker som en sterk kandidat. Eller motsatt: en sterk kandidat skal trekkes like mye for mangler som en svak kandidat. Dersom en sensor har gjort seg opp en mening om hva konklusjonen vil bli, er det alltid en fare for at vedkommende kun fokuserer på informasjon egnet til bekrefte eget standpunkt, og overser informasjon som peker i andre retninger (Eisner, 2017, s. 54).

Utdanningsdirektoratets Eksamensgruppe peker på sentrale begreper som kjennetegner kvalitet i eksamenssystemet, og som skal tilstrebes i alt vurderingsarbeid. Blant disse er validitet, reliabilitet og rettferdighet (Eksamensgruppa, 2019). Konstruktvaliditeten er en av de fremste begrunnelsene for autentisk vurdering (Gulikers et al., 2004, s. 68). Kan i det hele tatt kravene til rettferdighet, reliabilitet og validitet ivaretas i et komplekst emne som etterforskning, all den tid vurdererne er subjektive og kan vektlegge mange ulike ting på ulike nivå? Mine funn knyttet til subjektivitet, ulik bruk av kriterier og standarder, og delvis

misforhold mellom det som vurderes og det oppgaven spør om, vil selvsagt utfordre både reliabiliteten og validiteten i vurderingene. Vil da karakteren reflektere den kompetansen som arbeidsgiver tror at den gjør ved ansettelse? Spørsmålene er relevante både på individnivå og samfunnsnivå. Høyere utdanning er en av de største enkeltinvesteringer mennesker gjør, og arbeidslivet trenger pålitelig informasjon om hva studentene faktisk kan.

7. Oppsummering og veien videre

I dette prosjektet har jeg undersøkt hvordan interne sensorer ved PHS vurderer muntlige eksamensprestasjoner i emnet Etterforskning. Mine funn viser at sensorenes vurderinger av de samme prestasjoner er lite konsistente, og at interbedømmer reliabiliteten er lav. Vi vurderer forskjellig – i karakterbedømminger, spørsmålsstilling og begrunnelser – og bruken av pensum, karakterbeskrivelser, kriterier og standarder varierer. De eksplisitte kriteriene blir i liten grad henvist til i begrunnelsene, og sensorene synes ofte å støttes seg til implisitte kriterier basert på sitt profesjonelle skjønn og egen fagtilhørighet. Kommisjonskarakterene synes å bli satt et sted midt imellom kommisjonsmedlemmenes individuelle vurderinger, slik jeg også fant ved gjennomgang av tall fra klagesensur på skriftlig eksamen (kapittel 1.3). Da jeg startet arbeidet med prosjektet hadde jeg en antagelse om at vurderingene spriket, men jeg ble likevel overrasket over omfanget av ulikheter. I lys av kompleksiteten og omfanget av emnet – med fem fag, fem sensorveiledninger, 13 sider med eksplisitte vurderingskriterier, pensum og læringsutbytter – er det imidlertid ikke vanskelig å forstå spennet i ulikheter. Kandidaten får sagt mye i løpet av den halvtimen han eller hun har til disposisjon, og det sier seg selv at sensorene har stort spillerom for hva som kan vektlegges og verdsettes.

Funnene omkring lavt sensorsamsvar i karaktersettingen er ikke veldig oppsiktsvekkende, og har god støtte i eksisterende forskning om interbedømmer reliabilitet. Jeg har funnet lite forskning på muntlig eksamen, karakterbegrunnelser og på hvordan ulike vurderinger vektlegges i samspillet når karakter skal settes i kommisjoner, og gjorde flere oppdagelser i empirien rundt slike forhold. Blant annet fant jeg at forhold ved argumentasjonen, herunder dens innhold, kraft og rekkefølge, har betydning for karaktersettingen. Jeg oppdaget at det ikke utelukkende er faglige argumenter som synes å ha overtalelseskraft i karakterdiskusjoner, og at karakterdiskusjonens innledning legger føringer for det videre fokuset i samtalen. Jeg fant også at begrunnelsene for karakterbedømmingen tidvis kjennetegnes av å være lite

utdypet, og at det derfor kan være vanskelig å forstå hva som konkret ligger i dem. Jeg hadde ingen forventninger eller tanker knyttet til disse funnene på forhånd.

Det er et uunngåelig faktum at skjønn og subjektivitet alltid vil gjøre seg gjeldende innen kompetansebasert og autentisk vurdering. Spørsmålet er imidlertid i hvor stor grad det profesjonelle skjønn, herunder personlige krav, standarder og verdier, skal få dominere, all den tid det går på bekostning av målsetningene om at vurderingene skal være rettferdige og upartiske på en måte som sikrer utdanningskvaliteten (Meld. St. 16 (2016-2017); NOU 2020:3; UH-loven, 2005). Eller sagt med ordene til Bjølseth et al. (2011, s. 14): Kan vi leve med en situasjon der *hvem* som setter karakter kan ha (minst) like stor betydning som *hva* som er prestert? I lys av kravene som stilles fra regjeringshold, er svaret på spørsmålet gitt. Hvordan kan vi da bevege oss i retning av økt sensorsamsvar?

Bjølseth et al. (2011) brukte et år på å nærmest tvinge sensorer ved en profesjonsutdanning til å uttrykke sitt profesjonelle skjønn eksplisitt, og å gjøre dem til gjenstand for kollektiv analyse og drøfting – med formål å heve utdanningskvaliteten. Arbeidet bar imidlertid ikke frukter, og sensorsamsvaret ved profesjonsutdanningen fortsatte å være like lavt som tidligere. Likevel finnes det god støtte i forskning, som også påpekt av Bjølseth et al., om at utvikling av sensorkompetanse må skje gjennom refleksjon og samtale omkring læringsutbytter og kriterier, relatert til konkrete eksamensoppgaver (se for eksempel Baume & Yorke, 2002; Gynnild, 2010; O'Donovan, Price & Rust, 2001; Sadler, 2005). Også Gulikers et al. (2004) understreker viktigheten av sensorsamarbeid. Autentisitet er subjektivt, og oppfatninger av hva som er autentisk vil variere på tvers av lærere. Derfor er det avgjørende at lærerne snakker sammen om hvordan kriteriene skal forstås. Av samme grunn bør studentene inkluderes i vurderingsarbeidet. Selv om noen forskere hevder at man bør holde den summative vurderingen borte fra studentenes læringsprosess (se for eksempel Maugesten & Lauvås, 2004), er det god støtte i forskning til å hevde at studenter bør innlemmes i vurderingspraksis og ikke bare være objekter for andres vurdering (se for eksempel Elwood & Klenowski, 2002; Goos & Moni, 2001; Gynnild, 2010; Sadler, 2005). O'Donovan et al. har undersøkt studenters erfaringer med kriteriebasert vurdering ved et universitet i England, og funnet at kriteriene har liten praktisk betydning dersom de ikke forklares, eksemplifiseres og danner utgangspunkt for diskusjon og refleksjon. Slike prosesser vil også kunne bidra til økt felles forståelse omkring hva som kjennetegner ulike nivå av kriteriene (2001, s. 83). Dermed kan studentene rette studieinnsatsen mot relevante og meningsfulle læringsaktiviteter egnet til

å utvikle deres kompetanse i tråd med en felles forståelse av det de skal vurderes i.

For etterforskningsemnets del tror jeg vi vil ha mye å hente på tettere samarbeid mellom faglærerne om hvordan kriteriene skal forstås og hvordan de skal operasjonaliseres for studentene. Deretter bør den (mer) felles forståelsen inkluderes i undervisnings- og læringsopplegg, slik at studentene kan jobbe målrettet med aktiviteter egnet til å utvikle relevante og autentiske kompetanser i tråd med det som vil bli vektlagt ved vurderingen (Biggs, 1996; Gulikers et al., 2004). Man bør også diskutere kriterienes detaljnivå, fordi vurdering av komplekse kompetanser til syvende og sist bygger på en helhetlig opplevelse, som er noe annet enn summen av enkeltdeler. Selv om subjektivitet og skjønn alltid vil være tilstede og prege vurderingsarbeidet, vil slike grep kunne bidra til å holde sensorsamsvaret innen akseptable rammer og ivareta studentenes rettssikkerhet. Som påpekt i sensurprosjektet (Politihøgskolen, 2016b) kan grad av subjektivitet være særlig utfordrende for PHS, fordi utdanningen består av mange ulike fagtradisjoner. Representanter for ulike fag kan ha forskjellige oppfatninger av hva som kjennetegner kvalitet, og dermed verdsette og vektlegge vidt forskjellige forhold ved en prestasjon (Bjølseth et al., 2011; Friche, 2011). Derfor er et tett kollegialt samarbeid kanskje ekstra viktig for utvikling av felles forståelse og praksis.

Fra 2021 vil eksamensformen i etterforskningsemnet bli endret fra muntlig til skriftlig. Endringen er begrunnet i administrative forhold. Dermed mister man muligheten til å måle visse kompetanser, for eksempel det å presentere et produkt muntlig. En mulig fordel kan være at et skriftlig produkt er etterprøvbart, og at studentene kan klage på vurderingen. Imidlertid er det ingenting i veien for å også dokumentere en muntlig prestasjon – dersom man ønsker det. For øvrig kan jeg ikke se at endring av eksamensform vil løse utfordringene knyttet til vurdering og sensorsamsvar beskrevet i denne oppgaven. Min analyse innledningsvis, hvor jeg gjennomgikk tall fra klagesensur i tre fag ved PHS høsten 2018, illustrerer store utfordringer knyttet til interbedømmer reliabilitet også ved skriftlig eksamen. Ved denne eksamensformen vil trolig også antall sensorer per oppgave reduseres, slik at det blir færre blikk på samme produkt.

For fremtiden bør vurderingsordningen i etterforskningsemnet grundig evalueres, og det bør kanskje tenkes helt nytt rundt hvordan studentenes kompetanser skal prøves. Én test er ikke tilstrekkelig til å kunne fatte gyldige og rettfærdige slutninger omkring en students kompetanse, og kompetansebedømming bør derfor baseres på flere vurderinger ved bruk av

ulike vurderingsformer i ulike kontekster (Baartman et al., 2007, s. 114; Darling-Hammond & Snyder, 2000, s. 527). I en studie gjennomført ved Høgskolen i Østfold (HiØ) fant man at mappevurdering ga bedre resultater (karakterer) enn tradisjonell skriftlig eksamen (Kristiansen, Nätt & Heide, 2013). Det ble pekt på flere mulige årsaker til funnene, herunder at mappevurdering bedre tilrettelegger for læring og forståelse, og at studentene ved denne formen gis større anledning til å vise hva de kan. Også betydningen av tilbakemelding underveis ble trukket frem som en mulig årsak, både som motivasjonsfaktor og som veiviser for videre arbeid (s. 79). Et alternativ til mappevurdering, er å se hen til medisinutdanningen og deres utvikling av Objective Structured Clinical Examination (OSCE) – som er en form for stasjonseksamen bestående av ulike oppgaver med ulikt fokus, egnet til å teste et bredt spekter av kompetanser. Tilnærmingen er basert på forskning i medisinfaget som viser at et egnet utvalg eksamensoppgaver av ulik kontekst og type, som vurderes av ulike sensorer, kan sikre høy reliabilitet (Norcini et al., 2018; Wass, Van Der Vleuten, Shatzer & Jones, 2001).

I dag synes det å eksistere en form for forståelse i akademiske kretser om at vurderingene som gjøres av studentprestasjoner er riktige, og at kvalifiserte bedømmere vurderer et arbeid tilnærmet likt (Raaheim, 2000, s. 204). Karakterene brukes også i stor grad i tråd med en slik forståelse, ved at de fungerer som seleksjonsfiltre ved opptak til videre studier og ved ansettelse. Mitt studie er ett av flere som viser at vurderinger gjøres ulikt. Fremover vil det være nødvendig å få dypere innsikt i faktorer egnet til å utvikle vurderingsordninger som sikrer god og riktig læring og rettfærdig sensur. Det er behov for å rette større fokus mot kvaliteten i vurderingsarbeidet, slik at vurderingene – i tråd med overordnede målsetninger og forventninger – gir et pålitelig og gyldig bilde av studentenes læring og kompetanser.

Litteraturliste

- Aadland, I. J. (2019, 19. mars). Tal frå utdanningsinstitusjonane viser: Ja, det løner seg å klage på karakterane! *Framtida.no*. Hentet fra <https://framtida.no/2019/02/26/loner-det-seg-a-klage-pa-karakterane-ja>
- Andersen, Ø. (1995). *I retorikkens hage*. Oslo: Universitetsforl.
- Arter, J. A. & Spandel, V. (1992). Using Portfolios of Student Work in Instruction and Assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-44. <https://doi.org/10.1111/j.1745-3992.1992.tb00230.x>
- Asmyhr, M. (2011). Om vurdering av essaybesvarelser i høyere utdanning – en studie av vurderer-reliabilitet. *Uniped*, 34(04), 17-33.
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A. & van Der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2(2), 114-129. <https://doi.org/10.1016/j.edurev.2007.06.001>
- Baume, D. & Yorke, M. (2002, 2002/02/01). The Reliability of Assessment by Portfolio on a Course to Develop and Accredite Teachers in Higher Education. *Studies in Higher Education*, 27(1), 7-25. <https://doi.org/10.1080/03075070120099340>
- Berger, P. L. & Luckmann, T. (1966). *The social construction of reality: A Treatise in the Sociology of Knowledge*. Garden City: Anchor Books.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364.
- Bjølseth, G., Havnes, A. & Lauvås, P. (2011). Lavt sensorsamsvar – kan det bedres? *Uniped*, 34(04), 4-16.
- Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education*, 34(2), 209-220. <https://doi.org/10.1080/02602930801955978>
- Burner, T. & Berge, K. L. (2020, 3. april). Eksamen svekker mulighetene for god læringskultur, Helgesen. *Khrono*. Hentet fra <https://khrono.no/eksamen-svekker-mulighetene-for-god-laeringskultur-helgesen/477044>
- Charmaz, K. (2014). *Constructing grounded theory: A Practical Guide Through Qualitative Analysis* (2. utg.). London: Sage.
- Dalland, O. (2012). *Metode og oppgaveskriving* (5. utg.). Oslo: Gyldendal Akademisk.
- Darling-Hammond, L. & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5), 523-545. [https://doi.org/10.1016/S0742-051X\(00\)00015-9](https://doi.org/10.1016/S0742-051X(00)00015-9)
- Det norske universitetsråd. (2000). *Mål med mening: En utredning om etablering av en felles nasjonal karakterskala*. Agder, Bergen, Oslo, Tromsø, Trondheim.
- Dobson, S. (2007). Teoretisering rundt muntlig eksamen - en kvalitativ tilnærming. *Norsk Pedagogisk Tidsskrift*, 91(02), 137-149.
- Dreyfus, H. & Dreyfus, S. (2012). Fem stadier af færdighedstilegnelse: Fra nybegynder til ekspert. I K. Illeris (Red.), *49 tekster om læring* (s. 423-436). Frederiksberg: Samfundslitteratur.
- EC. (2007). *Key Competencies for Lifelong Learning: European Reference Framework*. Luxemburg: Office for Official Publications of the European Communities.
- Eide, T. (1999). *Retorisk leksikon*. Oslo: Spartacus.
- Eide, T. (2006). *Aristoteles Retorikk* (T. Eide, Overs.). Oslo: Vidarforl.
- Eisner, E. W. (2005). *Reimagining Schools - The Selected Works of Elliot W. Eisner* (1. utg.). London: Routledge.
- Eisner, E. W. (2017). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Teachers College Press.

- Eksamensgruppa. (2019). *Kunnskapsgrunnlag for evaluering av eksamensordningen*. Utdanningsdirektoratet. Hentet fra <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/Kunnskapsgrunnlag-for-evaluering-av-eksamensordningen/>
- Elwood, J. & Klenowski, V. (2002). Creating Communities of Shared Practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, 27(3), 243-256. <https://doi.org/10.1080/02602930220138606>
- Fautley, M. (2010). *Assessment in Music Education*. I. Oxford: Oxford University Press.
- FEK. (2016). *Generelle forskningsetiske retningslinjer* De nasjonale forskningsetiske komiteene. Hentet fra https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/fek_generelle_retningslinjer.pdf
- Forskrift om studier og eksamen ved Politihøgskolen. (2017). Hentet fra <https://lovdata.no/dokument/LTI/forskrift/2016-12-07-1957>
- Friche, N. (2011). Logikker for evaluering i erhvervsuddannelser. I K. Andreasen, N. Friche & A. Rasmussen (Red.), *Målt & Vejet: Uddannelsesforskning om evaluering* (s. 211-233). Aalborg: Aalborg Universitetsforlag.
- Gibbs, G. (2006). How assessment frames student learning. I C. Bryan & K. Clegg (Red.), *Innovative Assessment in Higher Education* (s. 23-36). London: Routledge.
- Gibbs, G. (2018). *Analyzing qualitative data* (2. utg.). Los Angeles: SAGE.
- Gielen, S., Dochy, F. & Dierick, S. (2003). The influence of assessment on learning. I M. Segers, F. Dochy & E. Cascallar (Red.), *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Goos, M. & Moni, K. (2001). Modelling Professional Practice: A collaborative approach to developing criteria and standards-based assessment in pre-service teacher education courses. *Assessment & Evaluation in Higher Education*, 26(1), 73-88. <https://doi.org/10.1080/02602930020022291a>
- Grønmo, S. (2004). *Samfunnsvitenskapelige metoder*. Bergen: Fagbokforl.
- Grønmo, S. (2012). Kvalitative og kvantitative metoder: Begreper og distinksjoner. *Sosiologisk tidsskrift*, 20(1), 85-91.
- Gulikers, J., Bastiaens, T. & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86.
- Gynnild, V. (2010). *Kriterier og skjønn i evaluering: En kasusstudie i utøvende musikkutdanning*. NOKUT.
- Gynnild, V. (2013). «Kriteriebasert vurdering» - hva innebærer det i praksis? *Uniped*, (01), 26-40.
- Hagen, A. (2016, 13. juli). Brit Ailen klagde på C-karakter: Endte opp med stryk. *NRK.no*. Hentet fra <https://www.nrk.no/norge/brit-ailen-klagde-pa-c-karakter-endte-opp-med-stryk-1.13040410>
- Halkier, B. (2003). The Challenge of Qualitative Generalisations in Communication Research. *Nordicom Review*, 24(1), 115-124. <https://doi.org/10.1515/nor-2017-0303>
- Halkier, B. (2012). Fokusgrupper. I S. Brinkmann & L. Tanggaard (Red.), *Kvalitative metoder: Empiri og teoriutvikling* (s. 133-152). Oslo: Gyldendal akademisk.
- Halvorsen, K. (2008). *Å forske på samfunnet: En innføring i samfunnsvitenskapelig metode* (5. utg.). Oslo: Cappelen Akademisk Forlag.
- Harlen, W. (2005). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245-270. <https://doi.org/10.1080/02671520500193744>
- Inderhaug, E. (2019, 28. september). Frykter at samholdet slår sprekker. *Politiforum*. Hentet fra <https://www.politiforum.no/artikler/frykter-at-samholdet-slar-sprekker/475135>
- Johannessen, L. E. F., Rafoss, T. W. & Rasmussen, E. B. (2018). *Hvordan bruke teori? Nyttige verktøy i kvalitativ analyse*. Oslo: Universitetsforlaget.
- Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Kjeldsen, J. E. (2002). *Visuel retorik* Institutt for medievitenskap, Universitetet i Bergen, Bergen.

- Kristiansen, M., Nätt, T. H. & Heide, C. F. (2013). Kvantitativ undersøkelse av mulige sammenhenger mellom vurderingsform og karakterer i høyere utdanning. *Uniped*, 36(2), 62-80.
- Kunnskapsdepartementet. (2014). *Nasjonalt kvalifikasjonsrammeverk for livslang læring*. Kunnskapsdepartementet. Hentet fra <https://www.regjeringen.no/contentassets/e579f913fa1d45c2bf2219afc726670b/nkr.pdf>
- Kvale, S. & Brinkmann, S. (2015). *Det kvalitative forskningsintervju* (3. utg., T. M. Anderssen & J. Ruge, Overs.). Oslo: Gyldendal Akademisk.
- Lizzio, A. & Wilson, K. (2004). Action Learning in Higher Education: An investigation of its potential to develop professional capability. *Studies in Higher Education*, 29(4), 469-488. <https://doi.org/10.1080/0307507042000236371>
- Mathy, J. (2018, 21. mars). Student klaget og gikk fra B til F. *Universitas.no*. Hentet fra <https://universitas.no/sak/64012/student-klaget-og-gikk-fra-b-til-f/>
- Maugesten, M. & Lauvås, P. (2004). *Bedre læring av matematikk ved enkle midler: Rapport fra et utviklingsprosjekt* (2004:6). Halden: Høgskolen i Østfold.
- Meld. St. 16 (2015-2016). (2016). *Fra utenforskap til ny sjanse. Samordnet innsats for voksnes læring*. Hentet fra <https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20152016/id2476199/>
- Meld. St. 16 (2016-2017). (2017). *Kultur for kvalitet i høyere utdanning*. Hentet fra <https://www.regjeringen.no/contentassets/aee30e4b7d3241d5bd89db69fe38f7ba/no/pdfs/stm201620170016000dddpdfs.pdf>
- Mellingsæter, H., Torset, N. S. & Husøy, E. (2018, 17. august). 200.000 eksamensvurderinger analysert: I nesten 15.000 tilfeller var spriket mellom sensorene to karakterer eller mer. *Aftenposten*. Hentet fra <https://www.aftenposten.no/norge/i/EoAplA/200000-eksamensvurderinger-analysert-i-nesten-15000-tilfeller-var-spraket-mellom-sensorene-to-karakterer-eller-mer>
- Messick, S. (1987). Validity. *Educational Testing Service*, 1987(2), i-208. <https://doi.org/10.1002/j.2330-8516.1987.tb00244.x>
- Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher*, 23(2), 13-23. <https://doi.org/10.3102/0013189X023002013>
- Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. *American Psychologist*, 50(9), 741. <https://doi.org/10.1037/0003-066X.50.9.741>
- Moss, P. A., Girard, B. J. & Haniford, L. C. (2006). Validity in Educational Assessment. *Review of Research in Education*, 30(1), 109-162. Hentet fra <https://doi.org/10.3102/0091732X030001109>
- Mulder, M. (2014). Conceptions of Professional Competence. I S. Billett, C. Harteis & H. Gruber (Red.), *International Handbook of Research in Professional and Practice-based Learning* (s. 107-137). Dordrecht: Springer.
- NESH. (2016). *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi* (4. utg. utg.) Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH). Hentet fra https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/60125_fek_retningslinjer_nesh_digital.pdf
- Norcini, J., Anderson, M. B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., ... Swanson, D. (2018, 2018/11/02). 2018 Consensus framework for good assessment. *Medical Teacher*, 40(11), 1102-1109. <https://doi.org/10.1080/0142159X.2018.1500016>
- Nore, H. & Lahn, L. C. (2015). Resultatkvalitet uttrykt gjennom fag- og svenneprøver. I H. Høst (Red.), *Kvalitet i fag- og yrkesopplæringen: Sluttrapport* (bd. NIFU-rapport 14/2015, Fafo-rapport 2015:32, HiOA rapport 2015 nr 4, s. 145-171). Oslo: Norsk institutt for studier av innovasjon, forskning og utdanning.
- NOU 2018:2. (2018). *Fremtidige kompetansebehov I - Kunnskapsgrunnlaget*. Kunnskapsdepartementet. Hentet fra <https://www.regjeringen.no/no/dokumenter/nou-2018-2/id2588070/>

- NOU 2020:2. (2020). *Fremtidige kompetansebehov III - Læring og kompetanse i alle ledd*. Kunnskapsdepartementet. Hentet fra <https://www.regjeringen.no/no/dokumenter/nou-2020-2/id2689744/>
- NOU 2020:3. (2020). *Ny lov om universiteter og høyskoler*. Hentet fra <https://www.regjeringen.no/contentassets/65a475004027495a8009b359e253b19e/no/pdfs/nou202020200003000dddpdfs.pdf>
- O'Donovan, B., Price, M. & Rust, C. (2001). The Student Experience of Criterion-Referenced Assessment (Through the Introduction of a Common Criteria Assessment Grid). *Innovations in Education and Teaching International*, 38(1), 74-85. <https://doi.org/10.1080/147032901300002873>
- Polithøgskolen. (2014). *Rammeplan for Bachelor - politiutdanning*. Oslo: Polithøgskolen.
- Polithøgskolen. (2015). *Fagplan Bachelor - politiutdanning 2015-2018*. Oslo: Polithøgskolen.
- Polithøgskolen. (2016a). *Kvalitetssystemet - Beskrivelse av Polithøgskolens systematiske kvalitetsarbeid*. Oslo: Polithøgskolen.
- Polithøgskolen. (2016b). *Sensurprosjektet - Rapport*. Oslo: Polithøgskolen.
- Polithøgskolen. (2019a). *Kvalitetsrapport studieår 2018-2019, Studieprogram Bachelor - Politiutdanning, Alle studiesteder*. Oslo: Polithøgskolen.
- Polithøgskolen. (2019b). *Rammeplan for Bachelor - politiutdanning 2019-2022*. Polithøgskolen. Hentet fra <https://www.politihogskolen.no/bachelor-politiutdanning/>
- Postholm, M. B. (2010). *Kvalitativ metode : en innføring med fokus på fenomenologi, etnografi og kasusstudier* (2. utg.). Oslo: Universitetsforlaget.
- Raaheim, A. (2000). En studie av inter-bedømmer reliabilitet ved eksamen på psykologi grunnfag. *Tidsskrift for den norske psykologiforening*, 37(03), 203-213.
- Raaheim, A. (2016). *Eksamensrevolusjonen: Råd og tips om eksamen og alternative vurderingsformer*. Oslo: Gyldendal akademisk.
- Rye, J. F. (2014). Konsistente karakterer? En undersøkelse av sensorreliabilitet i sosiologifaget. *Uniped*, 37(03), 63-77. <https://doi.org/10.3402/uniped.v37.22654>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *An International Journal of Learning and Cognition*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>
- Sadler, D. R. (2005). Interpretations of criteria - based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175-194. <https://doi.org/10.1080/0260293042000264262>
- Sadler, D. R. (2009a). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34(7), 807-826. <https://doi.org/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.
- Schuwirth, L. W. T. & Van Der Vleuten, C. P. M. (2004). Changing education, changing assessment, changing research? *Medical Education*, 38(8), 805-812. <https://doi.org/10.1111/j.1365-2929.2004.01851.x>
- Slaattelid, H. (2006). Retorisk skattkammer. *Nytt Norsk Tidsskrift*, (04), 386-391.
- St.meld. nr. 27 (2000-2001). (2001). *Gjør din plikt - Krev din rett. Kvalitetsreform av høyere utdanning*. Hentet fra <https://www.regjeringen.no/contentassets/eebf61fb4a204feb84e33355f30ad1a1/no/pdfa/stm200020010027000dddpdfa.pdf>
- Strøm, B., Falch, T., Gunnes, T. & Haraldsvik, M. (2013). *Karakterbruk i høyere utdanning (SØF-prosjekt nr. 2150: Analyse av karakterbruk og kvalitet i høyere utdanning)*. Hentet fra https://www.regjeringen.no/globalassets/upload/kd/karakterbruk_og_kvalitet_i_hoyere_utdanning.pdf
- Studietilsynsforskriften. (2017). *Forskrift om tilsyn med utdanningskvaliteten i høyere utdanning av 09.02.2017*. Hentet fra <https://lovdata.no/dokument/SF/forskrift/2017-02-07-137>
- Svendsen, L. F. H. & Grue, J. (2019, 15. mai). retorikk. Hentet 8. januar 2020 fra <https://snl.no/retorikk>

- Säljö, R. (2001). *Läring i praksis: Et sosiokulturelt perspektiv*. Oslo: Cappelen Akademisk.
- Tjora, A. H. (2017). *Kvalitative forskningsmetoder i praksis* (3. utg.). Oslo: Gyldendal Akademisk.
- UH-loven. (2005). *Lov om universiteter og høyskoler av 01.08.2005*. Hentet fra <https://lovdata.no/dokument/NL/lov/2005-04-01-15?q=lov> om universiteter og høyskoler
- Universitets- og høgskolerådet. (2008). *Generelle karakterbeskrivelser for UH-sektoren: Rapport fra en arbeidsgruppe oppnevnt av Universitets- og høgskolerådet*. Oslo: Universitets- og høgskolerådet.
- Universitets- og høgskolerådet. (2011). *Karaktersystemet - generelle, kvalitative beskrivelser*. Hentet fra https://www.uhr.no/f/p1/i4bfb251a-5e7c-4e34-916b-85478c61a800/karaktersystemet_generelle_kvalitative_beskrivelser.pdf
- Utdannings- og forskningsdepartementet. (2002). *Om lov om endringer i lov 12. mai 1995 nr. 22 om universiteter og høyskoler og lov 2. juli 1999 nr. 64 om helsepersonell* (Ot.prp. nr. 40 (2001-2002)). Hentet fra <https://www.regjeringen.no/contentassets/af788056c519440b8455ee833fdcae1f/no/pdfa/ot2001200220040000dddpdfa.pdf>
- Utdannings- og forskningsdepartementet. (2004). *Retningslinjer for bruk av det nasjonale karaktersystemet*. Hentet fra <https://www.regjeringen.no/no/dokumenter/retningslinjer-for-bruk-at-det-nasjonale/id91189/>
- Van Der Vleuten, C. P. M. & Schuwirth, L. W. T. (2005). Assessing professional competence: from methods to programmes. *Medical Education*, 39(3), 309-317. <https://doi.org/10.1111/j.1365-2929.2005.02094.x>
- Vinge, J. (2012). Analytiske og holistiske tilganger til vurdering – en teoretisk drøfting av vurderingsstrategier og vurderingshandlinger i musikkfaget. *Nordisk musikkpedagogisk forskning, Årbok 13*, 199-220. Hentet fra <http://hdl.handle.net/11250/172359>
- Wass, V., Van Der Vleuten, C., Shatzer, J. & Jones, R. (2001). Assessment of clinical competence. *The Lancet*, 357(9260), 945-949. [https://doi.org/10.1016/S0140-6736\(00\)04221-5](https://doi.org/10.1016/S0140-6736(00)04221-5)
- Wiggins, G. (1989). Teaching to the (Authentic) Test. *Educational Leadership*, 46(7), 41-47.
- Wilkinson, S. (2004). Focus group research. I D. Silverman (Red.), *Qualitative research: Theory, method and practice* (s. 177-199). London: Sage.
- Ørjasæter, E. (2016, 9. september). Sensurlotto. *Morgenbladet*. Hentet fra <https://morgenbladet.no/pafyll/2016/09/elin-orjasaeter-sensurlotto>

Vedlegg

VEDLEGG 1: MELDESKJEMA TIL NSD – VIDEOOPPTAK AV MUNTlig EKSAMEN.....	91
VEDLEGG 2: GODKJENNING FRA NSD – VIDEOOPPTAK AV MUNTlig EKSAMEN.....	97
VEDLEGG 3: INFORMASJONSSKRIV – SAMTYKKE TIL VIDEOOPPTAK AV MUNTlig EKSAMEN – TIL STUDENTER	98
VEDLEGG 4: INFORMASJONSSKRIV – SAMTYKKE TIL VIDEOOPPTAK AV MUNTlig EKSAMEN – TIL KOMMISJONSMEDLEMMER.....	99
VEDLEGG 5: PROSJEKTBEskRIVELSE.....	100
VEDLEGG 6: MELDESKJEMA TIL NSD - HELE PROSJEKTET	106
VEDLEGG 7: MELDINGSDIALOG MED NSD	112
VEDLEGG 8: GODKJENNING FRA NSD – HELE PROSJEKTET	115
VEDLEGG 9: INFORMASJONSSKRIV – DELTAKELSE I FORSKNINGSPROSJEKT – TIL SENSORER	117
VEDLEGG 10: SPØRRRESKJEMA – DEL 1	120
VEDLEGG 11: SPØRRRESKJEMA – DEL 2	121
VEDLEGG 12: INTERVJUGUIDE – FOKUSGRUPPEINTERVJU.....	127
VEDLEGG 13: MAIL TIL POTENSIELLE INFORMANTER – DELTAKELSE I FORSKNINGSPROSJEKT	128
VEDLEGG 14: MAIL TIL INFORMANTER I FORKANT AV DELTAKELSE (INDIVIDUELL OG FOKUSGRUPPE)	129
VEDLEGG 15: KODEBOK – ANALYSEKATEGORIER	130



MELDESKJEMA

Meldeskjema (versjon 1.6) for forsknings- og studentprosjekt som medfører meldeplikt eller konsesjonsplikt (jf. personopplysningsloven og helseregisterloven med forskrifter).

1. Intro		
Samles det inn direkte personidentifiserende opplysninger?	Ja • Nei ○	En person vil være direkte identifiserbar via navn, personnummer, eller andre personentydige kjennetegn. Les mer om hva personopplysninger er.
Hvis ja, hvilke?	<input checked="" type="checkbox"/> Navn <input type="checkbox"/> 11-sifret fødselsnummer <input type="checkbox"/> Adresse <input checked="" type="checkbox"/> E-post <input checked="" type="checkbox"/> Telefonnummer <input checked="" type="checkbox"/> Annet	NB! Selv om opplysningene skal anonymiseres i oppgave/rapport, må det krysses av dersom det skal innhentes/registreres personidentifiserende opplysninger i forbindelse med prosjektet. Les mer om hva behandling av personopplysninger innebærer.
Annet, spesifiser hvilke	Videopptak av personen (muntlig eksamensgjennomføring). Det er ikke sikkert at telefonnummer blir samlet inn, det avhenger av behov for å kontakte studentene i forkant for avtale (eller om dette gjøres per mail).	
Skal direkte personidentifiserende opplysninger kobles til datamaterialet (koblingsnøkkel)?	Ja • Nei ○	Merk at meldeplikten utløses selv om du ikke får tilgang til koblingsnøkkel, slik fremgangsmåten ofte er når man benytter en databehandler.
Samles det inn bakgrunnsopplysninger som kan identifisere enkeltpersoner (indirekte personidentifiserende opplysninger)?	Ja • Nei ○	En person vil være indirekte identifiserbar dersom det er mulig å identifisere vedkommende gjennom bakgrunnsopplysninger som for eksempel bostedskommune eller arbeidsplass/skole kombinert med opplysninger som alder, kjønn, yrke, diagnose, etc.
Hvis ja, hvilke	Videopptak av personen (muntlig eksamensgjennomføring)	NB! For at stemme skal regnes som personidentifiserende, må denne bli registrert i kombinasjon med andre opplysninger, slik at personer kan gjenkjennes.
Skal det registreres personopplysninger (direkte/indirekte/via IP-lepost adresse, etc) ved hjelp av nettbaserte sporeskjema?	Ja ○ Nei •	Les mer om nettbaserte spareskjema .
Blir det registrert personopplysninger på digitale bilde- eller videopptak?	Ja • Nei ○	Bilde/videopptak av ansikter vil regnes som personidentifiserende.
Søkes det vurdering fra REK om hvorvidt prosjektet er omfattet av helseforskningsloven?	Ja ○ Nei •	NB! Dersom REK (Regional Komité for medisinsk og helsefaglig forskningsetik) har vurdert prosjektet som helseforskning, er det ikke nødvendig å sende inn meldeskjema til personvernombudet (NB! Gjelder ikke prosjekter som skal benytte data fra pseudonyme helseregistre). Les mer. Dersom tilbakemelding fra REK ikke foreligger, anbefaler vi at du avventer videre utfylling til svar fra REK foreligger.
2. Prosjekttittel		
Prosjekttittel	Howdan vurderes muntlig eksamen i emnet etterforskning ved Polithøgskolen?	Oppgi prosjektets tittel. NB! Dette kan ikke være «Masteroppgave» eller tilknyttet, navnet må beskrive prosjektets innhold.
3. Behandlingsansvarlig institusjon		
Institusjon	OsloMet - Storbyuniversitetet	Velg den institusjonen du er tilknyttet. Alle nivå må oppgis. Ved studentprosjekt er det studentens tilknytning som er avgjørende. Dersom institusjonen ikke finnes på listen, har den ikke avtale med NSD som personvernombud. Vennligst ta kontakt med institusjonen. Les mer om behandlingsansvarlig institusjon .
Avdeling/Fakultet	Fakultet for lærerutdanning og internasjonale studier	
Institutt	Institutt for yrkesfaglærerutdanning	
4. Daglig ansvarlig (forsker, veileder, stipendiat)		

Fornavn	Hæge	<p>Før opp navnet på den som har det daglige ansvaret for prosjektet. Veileder er vanligvis daglig ansvarlig ved studentprosjekt. Les mer om daglig ansvarlig.</p> <p>Daglig ansvarlig og student må i utgangspunktet være tilknyttet samme institusjon. Dersom studenten har ekstern veileder, kan biveileder eller fagansvarlig ved studiestedet stå som daglig ansvarlig.</p> <p>Arbeidssted må være tilknyttet behandlingsansvarlig institusjon, f.eks. underavdeling, institutt etc.</p> <p>NB! Det er viktig at du oppgir en e-postadresse som brukes aktivt. Vennligst gi oss beskjed dersom den endres.</p>
Etternavn	Nore	
Stilling	Førstelektor	
Telefon	92442909	
Mobil		
E-post	hnore@oslomet.no	
Alternativ e-post	hnore@oslomet.no	
Arbeidssted	OsloMet	
Adresse (arb.)	Kunnskapsveien 55	
Postnr./sted (arb.sted)	2007 Kjeller	
5. Student (master, bachelor)		
Studentprosjekt	Ja ● Nei ○	Dersom det er flere studenter som samarbeider om et prosjekt, skal det velges en kontaktperson som føres opp her. Øvrige studenter kan føres opp under pkt 10.
Fornavn	Christine	
Etternavn	Nordsletten	
Telefon	41663269	
Mobil		
E-post	christinesatre@gmail.com	
Alternativ e-post	chrino@pfs.no	
Privatadresse	Frostveien 22	
Postnr./sted (privatadr.)	1164 Oslo	
Type oppgave	<ul style="list-style-type: none"> ● Masteroppgave ○ Bacheloroppgave ○ Semesteroppgave ○ Annet 	
6. Formålet med prosjektet		
Formål	Jeg ønsker å få dypere innsikt i hvordan ulike sensorer innen etterforskning vurderer eksamensprestasjoner, og hva som legges til grunn ved vurderingen. I de to ukene (muntlig) etterforskningseksamen gjennomføres er det mellom 40 og 50 studenter som eksamineres per dag fordelt på 10 kommisjoner. For å undersøke problemstillingen vil jeg filme to eksamensgjennomføringer og bruke dette som grunnlag for senere datainnsamling fra sensorer om hvordan de vurderer samme (video)prestasjon og med hvilken begrunnelse. Målet er å få mer kunnskap om vurderingene som gjøres av den enkelte, og å få grunnlag for å gjøre hensiktsmessige valg i det videre vurderingsarbeidet ved PolitiHøgskolen generelt og innen etterforskningsemet spesielt.	Redegjør kort for prosjektets formål, problemstilling, forsknings spørsmål e.l.
7. Hvilke personer skal det innhentes personopplysninger om (utvalg)?		
Kryss av for utvalg	<input type="checkbox"/> Bamehagebarn <input type="checkbox"/> Skoleelever <input type="checkbox"/> Pasienter <input type="checkbox"/> Brukere/klienter/kunder <input type="checkbox"/> Ansatte <input type="checkbox"/> Bamevemsbarn <input type="checkbox"/> Lærere <input type="checkbox"/> Helsepersonell <input type="checkbox"/> Asylsøkere <input checked="" type="checkbox"/> Andre	Les mer om forskjellige forskningstematikker og utvalg.

Beskriv utvalgdeltakere	På dette stadiet er det snakk om å filme to muntlige eksamensgjennomføringer. Dette må gjøres nå fordi det er et år til neste mulighet. Senere vil jeg bruke disse som grunnlag for å eksempelvis intervju sensorer i etterforskningsemnet for å undersøke deres vurderinger av de filmede prestasjonene. Det bemerkes at metodisk tilnærming for datainnsamling fra sensorene ikke er endelig bestemt. Dette må vurderes til høsten, og her vil jeg inkludere både veileder, kolleger og læringsgruppen.	Med utvalg menes dem som deltar i undersøkelsen eller dem det innhentes opplysninger om.
Rekruttering/trekking	Jeg har valgt studenter fra to klasser som skal opp til eksamen på tirsdag til uken (12.juni). Jeg skal da selv ikke være sensor (men er reserve), og har tid til å administrere datainnsamlingen ved å sette opp videokamera og mikrofon. Fra de klassene som skal opp denne dagen har jeg valgt fire gutter og fire jenter fra hver av klassene, totalt 16 studenter, som jeg planlegger å sende epost til. I eposten vil jeg vedlegge skriv om informert samtykke. Som nevnt trenger jeg kun samtykke fra to studenter, men på grunn av at datainnsamlingen skal gjøres allerede til uken velger jeg å kontakte flere i håp om å få positivt svar av minst to.	Beskriv hvordan utvalget trekkes eller rekrutteres og oppgi hvem som foretar den. Et utvalg kan rekrutteres gjennom f.eks. en bedrift, skole, idrettsmiljø eller eget nettverk, eller trekkes fra registre som f.eks. Folkeregisteret, SSB-registre, pasientregistre.
Førstegangskontakt	Jeg vil sende epost til aktuelle studenter	Beskriv hvordan førstegangskontakten opprettes og oppgi hvem som foretar den. Les mer om førstegangskontakt og forskjellige utvalg på våre temasider.
Alder på utvalget	<input type="checkbox"/> Barn (0-15 år) <input type="checkbox"/> Ungdom (16-17 år) <input checked="" type="checkbox"/> Voksne (over 18 år)	Les om forskning som involverer barn på våre nettsider.
Omtrentlig antall personer som inngår i utvalget	2 (16 får epost)	
Samles det inn sensitive personopplysninger?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	Les mer om sensitive opplysninger.
Hvis ja, hvilke?	<input type="checkbox"/> Rasemessig eller etnisk bakgrunn, eller politisk, filosofisk eller religiøs oppfatning <input type="checkbox"/> At en person har vært mistenkt, siktet, tiltalt eller dømt for en straffbar handling <input type="checkbox"/> Helseforhold <input type="checkbox"/> Seksuelle forhold <input type="checkbox"/> Medlemskap i fagforeninger	
Inkluderes det myndige personer med redusert eller manglende samtykkekompetanse?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	Les mer om pasienter, brukere og personer med redusert eller manglende samtykkekompetanse.
Samles det inn personopplysninger om personer som selv ikke deltar (tredjepersoner)?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	Med opplysninger om tredjeperson menes opplysninger som kan identifisere personer (direkte eller indirekte) som ikke inngår i utvalget. Eksempler på tredjeperson er kollega, elev, klient, famillemidlem, som identifiseres i datamaterialet. Les mer.
8. Metode for innsamling av personopplysninger		
Kryss av for hvilke datainnsamlingsmetoder og datakilder som vil benyttes	<input type="checkbox"/> Papirbasert spørreskjema <input type="checkbox"/> Elektronisk spørreskjema <input type="checkbox"/> Personlig intervju <input type="checkbox"/> Gruppeintervju <input type="checkbox"/> Observasjon <input type="checkbox"/> Deltakende observasjon <input type="checkbox"/> Blogg/sosiale medier/internett <input type="checkbox"/> Psykologiske/pedagogiske tester <input type="checkbox"/> Medisinske undersøkelser/tester <input type="checkbox"/> Journaldata (medisinske journaler)	<p>Personopplysninger kan innhentes direkte fra den registrerte f.eks. gjennom spørreskjema, intervju, tester, og/eller ulike journaler (f.eks. elevmapper, NAV, PPT, sykehus) og/eller registre (f.eks. Statistisk sentralbyrå, sentrale helseregistre).</p> <p>NB! Dersom personopplysninger innhentes fra forskjellige personer (utvalg) og med forskjellige metoder, må dette spesifiseres i kommentar-boksen. Husk også å legge ved relevante vedlegg til alle utvalgs-gruppene og metodene som skal benyttes.</p> <p>Les mer om registerstudier. Dersom du skal anvende registerdata, må variabeliste lastes opp under pkt. 15</p> <p>Les mer om forskningsmetoder.</p>
	<input type="checkbox"/> Registerdata	
	<input checked="" type="checkbox"/> Annen innsamlingsmetode	
Oppgi hvilken	Video	

Tilleggsopplysninger	De to videoopptakene må sikres nå fordi etterforskningseksamen kun gjennomføres denne og neste uke (i 2018). Arbeidet med masteroppgaven begynner først til høsten, og opptakene vil da danne grunnlaget for videre datainnsamling ved at de skal vises til utvalgte sensorer innen etterforskningseksamen. Sensorene vil deretter spørres om sine vurderinger i tilknytning til eksamensprestasjonen - altså få spørsmål egnet til å belyse problemstillingen. Denne datainnsamlingen (fra sensorer) vil bli gjennomført i løpet av høsten 2018 og/eller våren 2019. Jeg vil da melde inn mer detaljert informasjon om blant annet utvalg av disse og valg av metodisk tilnærming. Dette er ikke klart på nåværende tidspunkt.	
9. Informasjon og samtykke		
Oppgi hvordan utvalget/deltakerne informeres	<input checked="" type="checkbox"/> Skriftlig <input type="checkbox"/> Muntlig <input type="checkbox"/> Informeres ikke	<p>Dersom utvalget ikke skal informeres om behandlingen av personopplysninger må det begrunnes.</p> <p>Les mer.Vennligst send inn mal for skriftlig eller muntlig informasjon til deltakerne sammen med meldeskjema.</p> <p>Last ned en veiledende mal her.</p> <p>Les om krav til informasjon og samtykke.</p> <p>NB! Vedlegg lastes opp til sist i meldeskjemaet, se punkt 15 Vedlegg.</p>
Samtykker utvalget til deltakelse?	<input checked="" type="radio"/> Ja <input type="radio"/> Nei <input type="radio"/> Flere utvalg, ikke samtykke fra alle	<p>For at et samtykke til deltakelse i forskning skal være gyldig, må det være frivillig, uttrykkelig og informert.</p> <p>Samtykke kan gis skriftlig, muntlig eller gjennom en aktiv handling. For eksempel vil et besvart spørreskjema være å regne som et aktivt samtykke.</p> <p>Dersom det ikke skal innhentes samtykke, må det begrunnes. Les mer.</p>
10. Informasjonssikkerhet		
Hvordan oppbevares navnlisten/ koblingsnøkkelen og hvem har tilgang til den?	Den vil bli lagt på passordbeskyttet elektronisk plassering hvor kun jeg har tilgang	
Oppbevares direkte personidentifiserbare opplysninger på andre måter?	Ja <input checked="" type="radio"/> Nei <input type="radio"/>	
Spesifiser	Videoopptaket vil også bli lagret elektronisk, men på en annen plassering enn koblingsnøkkelen. Også denne plasseringen er passordbeskyttet og det vil kun være jeg som har tilgang.	NB! Som hovedregel bør ikke direkte personidentifiserende opplysninger registreres sammen med det øvrige datamaterialet. Vi anbefaler koblingsnøkkel .
Hvordan registreres og oppbevares personopplysningene?	<input type="checkbox"/> På server i virksomhetens nettverk <input type="checkbox"/> Fysisk isolert PC tilhørende virksomheten (dvs. ingen tilknytning til andre datamaskiner eller nettverk, interne eller eksterne) <input checked="" type="checkbox"/> Datamaskin i nettverkssystem tilknyttet Internett tilhørende virksomheten <input type="checkbox"/> Privat datamaskin <input type="checkbox"/> Videoopptak/fotografi <input type="checkbox"/> Lydopptak <input type="checkbox"/> Notater/papir <input checked="" type="checkbox"/> Mobile lagringsenheter (bærbar datamaskin, minnepenn, minnekort, cd, eksternt harddisk, mobiltelefon) <input type="checkbox"/> Annen registreringsmetode	<p>Merk av for hvilke hjelpemidler som benyttes for registrering og analyse av opplysninger.</p> <p>Sett flere kryss dersom opplysningene registreres på flere måter.</p> <p>Med «virksomhet» menes her behandlingsansvarlig institusjon.</p> <p>NB! Som hovedregel bør data som inneholder personopplysninger lagres på behandlingsansvarlig sin forskningsserver.</p> <p>Lagring på andre medier - som privat pc, mobiltelefon, minnepenne, server på annet arbeidssted - er mindre sikkert, og må derfor begrunnes. Slik lagring må avklares med behandlingsansvarlig institusjon, og personopplysningene bør krypteres.</p>
Annen registreringsmetode beskriv		
Hvordan er datamaterialet beskyttet mot at uvedkommende får innsyn?	Passordbeskyttet på PC (ett passord for å komme inn på selve PCen og et annet passord for å få tilgang til aktuell plassering.	Er f.eks. datamaskintilgangen beskyttet med brukernavn og passord, står datamaskinen i et låsbart rom, og hvordan sikres bærbare enheter, utskrift og opptak?
Samles opplysningene inn/behandles av en databehandler (ekstern aktør)?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	Dersom det benyttes eksterne til helt eller delvis å behandle personopplysninger, f.eks. Guestbook, transkriberingsassistent eller loik, er dette å betrakte som en databehandler . Slike oppdrag må kontraktreguleres.
Hvis ja, hvilken		

Overføres personopplysninger ved hjelp av e-post/Internett?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	F.eks. ved overføring av data til samarbeidspartner, databehandler mm.
Hvis ja, beskriv?		Dersom personopplysninger skal sendes via Internett, bør de krypteres tilstrekkelig. Vi anbefaler ikke lagring av personopplysninger på nettskytjenester. Bruk av nettskytjenester må avklares med behandlingsansvarlig institusjon. Dersom nettskytjeneste benyttes, skal det inngås skriftlig databehandleravtale med leverandøren av tjenesten. Les mer.
Skal andre personer enn daglig ansvarlig/student ha tilgang til datamaterialet med personopplysninger?	Ja <input checked="" type="radio"/> Nei <input type="radio"/>	
Hvis ja, hvem (oppgi navn og arbeidssted)?	Videooptaket vil bli vist til senere utvalg av sensorer. Det er datagrunnlaget som vil bli innhentet fra sensorene (etter å ha sett videoen) som vil belyse problemstillingen.	
Utliveres/deles personopplysninger med andre institusjoner eller land?	<input checked="" type="radio"/> Nei <input type="radio"/> Andre institusjoner <input type="radio"/> Institusjoner i andre land	F.eks. ved nasjonale samarbeidsprosjekter der personopplysninger utveksles eller ved internasjonale samarbeidsprosjekter der personopplysninger utveksles.
11. Vurdering/godkjenning fra andre instanser		
Søkes det om dispensasjon fra taushetsplikten for å få tilgang til data?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	For å få tilgang til taushetsbelagte opplysninger fra f.eks. NAV, PPT, sykehus, må det søkes om dispensasjon fra taushetsplikten . Dispensasjon søkes vanligvis fra aktuelt departement.
Hvis ja, hvilke		
Søkes det godkjenning fra andre instanser?	Ja <input type="radio"/> Nei <input checked="" type="radio"/>	I noen forskningsprosjekter kan det være nødvendig å søke flere tilatelser. Søkes det f.eks. om tilgang til data fra en registerer? Søkes det om tilatelse til forskning i en virksomhet eller en skole? Les mer om andre godkjenninger.
Hvis ja, hvilken		
12. Periode for behandling av personopplysninger		
Prosjektstart	04.06.2018	Prosjektstart Vennligst oppgi tidspunktet for når kontakt med utvalget skal gjøres/datainsamlingen starter.
Planlagt dato for avslutt	01.06.2020	Prosjektslutt: Vennligst oppgi tidspunktet for når datamaterialet enten skal anonymiseres/slettes, eller arkiveres i påvente av oppfølgingsstudier eller annet.
Skal personopplysninger publiseres (direkte eller indirekte)?	<input type="checkbox"/> Ja, direkte (navn e.l.) <input type="checkbox"/> Ja, indirekte (identifiserende bakgrunnsopplysninger) <input checked="" type="checkbox"/> Nei, publiseres anonymt	Les mer om direkte og indirekte personidentifiserende opplysninger. NB! Dersom personopplysninger skal publiseres, må det vanligvis innhentes eksplisitt samtykke til dette fra den enkelte, og deltakere bør gis anledning til å lese gjennom og godkjenne statser.
Hva skal stje med datamaterialet ved avslutt?	<input checked="" type="checkbox"/> Datamaterialet anonymiseres <input type="checkbox"/> Datamaterialet oppbevares med personidentifikasjon	NB! Her menes datamaterialet, ikke publikasjon. Selv om data publiseres med personidentifikasjon skal som regel øvrig data anonymiseres. Med anonymisering menes at datamaterialet bearbejdes slik at det ikke lenger er mulig å fere opplysningene tilbake til enkeltpersoner. Les mer om anonymisering av data.
13. Finansiering		
Hvordan finansieres prosjektet?		Fylles ut ved eventuell eksternt finansiering (oppdragsforskning, annet).
14. Tilleggsopplysninger		

Tilleggsopplysninger	<p>Jeg må som nevnt hente inn disse dataene i løpet av denne eller neste uke, fordi neste eksamensgjennomføring i etterforskningsmodulen skjer først om et år. Jeg har vært i kontakt med både egen leder, NSD (per telefon w/ Eva Paine) og min veileder ved OsloMet Høge Nore. Ved NSD ble jeg anbefalt å søke selvom det er tidlig i forløpet, og å bruke malen til NSD ved utforming av skriv for informert samtykke. Det har jeg gjort, og skrevet vil bli vedlagt mail som jeg skal sende til studentene som blir forespurt om å bli videofilmet ifm muntlig eksamensgjennomføring. OsloMet w/ Høge Nore er informert og inneforstått med prosjektet, at jeg melder det inn nå og at jeg samler inn (video)data til uken. De er trygge på at jeg behandler dataene med nødvendig forsiktighet i tråd med OsloMets retningslinjer. Jeg vil senere sende inn ytterligere informasjon knyttet til prosjektets fortsettelse, herunder utvalg og metodisk tilnærming for datainnsamling fra sensorer.</p>	<p>Dersom prosjektet er del av et prosjekt (eller skal ha data fra et prosjekt) som allerede har tilrømming fra personvernombudet og/eller konsesjon fra Datatilsynet, beskriv dette her og oppgi navn på prosjektleder, prosjektittel og/eller prosjektnummer.</p>
15. Vedlegg		
Vedlegg	<p>Antall vedlegg: 1.</p> <ul style="list-style-type: none"> • informasjonsskriv_videoopptak_av_muntlig_eksamensgjennomfoering.doc 	



Hæge Nore

2007 KJELLER

Vår dato: 04.07.2018

Vår ref: 61040 / 3 / TAL

Deres dato:

Deres ref:

Vurdering fra NSD Personvernombudet for forskning § 31

Personvernombudet for forskning viser til meldeskjema mottatt 07.06.2018 for prosjektet:

<i>61040</i>	<i>Hvordan vurderes muntlig eksamen i emnet etterforskning ved Politihøgskolen?</i>
<i>Behandlingsansvarlig</i>	<i>OsloMet - Storbyuniversitetet, ved institusjonens øverste leder</i>
<i>Daglig ansvarlig</i>	<i>Hæge Nore</i>
<i>Student</i>	<i>Christine Nordsletten</i>

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet og øvrig dokumentasjon finner vi at prosjektet er meldepliktig og at personopplysningene som blir samlet inn i dette prosjektet er regulert av personopplysningsloven § 31. På den neste siden er vår vurdering av prosjektopplegget slik det er meldt til oss. Du kan nå gå i gang med å behandle personopplysninger.

Vilkår for vår anbefaling

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

- opplysningene gitt i meldeskjemaet og øvrig dokumentasjon
- vår prosjektvurdering, se side 2
- eventuell korrespondanse med oss

Vi forutsetter at du ikke innhenter sensitive personopplysninger.

Meld fra hvis du gjør vesentlige endringer i prosjektet

Dersom prosjektet endrer seg, kan det være nødvendig å sende inn endringsmelding. På våre nettsider finner du svar på hvilke [endringer](#) du må melde, samt endringskjema.

Opplysninger om prosjektet blir lagt ut på våre nettsider og i Meldingsarkivet

Vi har lagt ut opplysninger om prosjektet på nettsidene våre. Alle våre institusjoner har også tilgang til egne prosjekter i [Meldingsarkivet](#).

Vi tar kontakt om status for behandling av personopplysninger ved prosjektslutt

Documentet er elektronisk produsert og godkjent ved NSDs nummer for elektronisk godkjenning.

Forespørsel om videoopptak av eksamensgjennomføring

”Hvordan vurderes muntlig eksamen i emnet Etterforskning ved Politihøgskolen?”

Ved å skrive under på dette skjemaet sier du ja til at din muntlige eksamensgjennomføring i emnet Etterforskning filmes. Videoopptaket vil bli brukt i forbindelse med et prosjekt hvor jeg vil undersøke hvordan etterforskningseksamen blir sensurert og vurdert.

Bakgrunn og formål

Undersøkelsen inngår i min masteroppgave ved masterstudium i yrkespedagogikk ved OsloMet. Etterforskningseksamen er på 18 studiepoeng, og en av de største som gjennomføres ved bachelorutdanningen ved Politihøgskolen (PHS). Jeg ønsker å få dypere innsikt i hvordan ulike sensorer innen etterforskning vurderer eksamensprestasjoner, og hva som legges til grunn ved vurderingen. Jeg ønsker å undersøke dette med utgangspunkt i reelle eksamensgjennomføringer (som er filmet). Jeg vil ta opp to eksamensgjennomføringer i forbindelse med prosjektet.

Hva innebærer deltakelse for deg?

Ditt bidrag innebærer at din eksamensgjennomføring filmes, og at denne brukes som grunnlag for videre undersøkelser blant sensorer.

Hva skjer med informasjonen om deg?

Alle dine personopplysninger vil bli behandlet konfidensielt, og både personopplysninger (navn og epostadresse) og videoopptak vil bli oppbevart av meg på passordbeskyttet plassering hvor ingen andre har tilgang. Personopplysningene vil lagres adskilt fra øvrige data. Videoopptaket vil kun bli vist til aktuelle sensorer (antall er per nå ikke avgjort). Du vil ikke bli omhandlet med navn eller andre personopplysninger i masteroppgaven, og vil ikke på noe vis kunne gjenkjennes. Det er sensorenes vurderinger som er det sentrale i undersøkelsen.

Arbeidet med masteroppgaven strekker seg fra høsten 2018 til våren 2020. Datainnsamlingen blant sensorene vil bli gjort i løpet av høsten 2018 og/eller våren 2019. Etter at eksamensresultatet foreligger vil dine personopplysninger anonymiseres og videoopptaket slettes.

Masteroppgaven vil være tilgjengelig for de som måtte ønske å lese den, særlig ved PHS, men også andre.

Frivillig deltakelse

Det er frivillig å la seg filme ved eksamensgjennomføringen, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med Christine Nordsletten på epost chrino@phs.no eller telefonnummer 41663269.

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Mvh Christine Nordsletten

Samtykke til videoopptak av muntlig eksamensgjennomføring

Jeg har mottatt informasjon om prosjektet, og samtykker i at min eksamensgjennomføring videofilmes og at opptaket brukes som grunnlag ved senere datainnsamling fra sensorer innen emnet.

Dato: Signatur:

Forespørsel om opptak av eksamensgjennomføring

”Hvordan vurderes muntlig eksamen i emnet Etterforskning ved Politihøgskolen?”

Ved å skrive under på dette skjemaet sier du ja til at det tas videoopptak av en eksamensgjennomføring i emnet Etterforskning hvor du er blant sensorene i kommisjonen. Videokameraet vil være rettet mot studenten og kommisjonen vil ikke bli filmet, men informasjonen som gis innledningsvis og spørsmålene du stiller avslutningsvis vil bli dokumentert med lydopptak. Det samme vil den etterfølgende vurderingsdiskusjonen blant sensorene vedrørende prestasjonen som foregår etter at kandidaten har forlatt rommet. Videoopptaket vil bli brukt i forbindelse med et prosjekt hvor jeg vil undersøke hvordan etterforskningseksamen blir sensurert og vurdert. Studenten som skal filmes har fått informasjon om prosjektet og har samtykket til videoopptak.

Bakgrunn og formål

Undersøkelsen inngår i min masteroppgave ved masterstudium i yrkespedagogikk ved OsloMet. Etterforskningseksamen er på 18 studiepoeng, og en av de største som gjennomføres ved bachelorutdanningen ved Politihøgskolen (PHS). Jeg ønsker å få dypere innsikt i hvordan ulike sensorer innen etterforskning vurderer eksamensprestasjoner, og hva som legges til grunn ved vurderingen. Jeg ønsker å undersøke dette med utgangspunkt i reelle eksamensgjennomføringer (som er filmet). Jeg vil ta opp to eksamensgjennomføringer i forbindelse med prosjektet.

Hva innebærer deltakelse for deg?

Ditt bidrag innebærer at en eksamensgjennomføring hvor du er en del av kommisjonen, blir videofilmet. Opptaket vil bli brukt som grunnlag for videre datainnsamling blant sensorer, ved at de får se opptaket og deretter svarer på spørsmål knyttet til hvordan de vurderer eksamensprestasjonen. Lydopptaket av kommisjonens diskusjoner/vurderinger som foregår etter at kandidaten har forlatt rommet, og som omhandler eksamensprestasjonen, vil kun bli hørt på av undertegnede og vil inngå som en del av datagrunnlaget som senere skal analyseres for å besvare problemstillingen. Det bemerkes at det per nå ikke er endelig bestemt hvorvidt vurderingene til kommisjonen skal inngå som en del av det senere datagrunnlaget. Dersom det besluttes at dette ikke skal brukes vil det umiddelbart bli slettet.

Hva skjer med informasjonen om deg?

Alle dine personopplysninger (i dette tilfelle ditt navn) vil bli behandlet konfidensielt, og både personopplysninger og videoopptak vil bli oppbevart av meg på passordbeskyttet plassering hvor ingen andre har tilgang. Personopplysningene vil lagres adskilt fra øvrige data. Videoopptaket vil kun bli vist til utvalgte sensorer (antall er per nå ikke avgjort). Det kan ikke utelukkes at noen av disse vil kjenne igjen din stemme fra lyden på opptaket. Du vil ikke bli omhandlet med navn eller andre personopplysninger i masteroppgaven, og vil ikke på noe vis kunne gjenkjennes.

Arbeidet med masteroppgaven strekker seg fra høsten 2018 til våren 2020. Datainnsamlingen blant sensorene vil bli gjort i løpet av høsten 2018 og/eller våren 2019. Etter at eksamensresultatet foreligger vil dine personopplysninger anonymiseres og opptaket slettes. Masteroppgaven vil være tilgjengelig for de som måtte ønske å lese den, særlig ved PHS, men også andre.

Frivillig deltakelse

Det er frivillig å delta i form av at det du sier under eksamensgjennomføringen og i den etterfølgende vurderingsdiskusjonen, blir tatt opp på lyd. Du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du har spørsmål til studien, ta kontakt med Christine Nordsletten på epost chrino@phs.no eller telefonnummer 41663269. Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Mvh Christine Nordsletten

Samtykke til opptak

Jeg har mottatt informasjon om prosjektet, og samtykker i at det tas videoopptak av kandidat hvor jeg sitter i kommisjonen, og at det jeg sier i løpet av den tiden kandidaten er i rommet og i den etterfølgende vurderingsdiskusjonen i kommisjonen, blir tatt opp på lyd.

Dato:

Signatur:

Christine Sætre Nordsletten

Prosjektbeskrivelse - MAYP5900

Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?

Innholdsfortegnelse

INNRAMMING	1
PROBLEMSTILLING OG FORSKNINGSPØRSMÅL	3
TEORETISK TILNÆRMING	3
METODISK TILNÆRMING	4
FREMDRIFTSPLAN	5
LITTERATURLISTE	6

Innramming

Denne oppgaven skal handle om *vurderernes vurderinger*. Jeg vil undersøke hvordan muntlig eksamen i etterforskningsemnet¹ ved Politihøgskolen (PHS) vurderes av interne sensorer.

Etterforskningsemnet består av fem fagelementer (jus, rapport- og etterforskningslære, psykologi, kriminalteknikk og digitalt politiarbeid). Eksamen er muntlig, og en av de mest omfattende i studiet ved at den utgjør totalt 18 studiepoeng. Studentene skal redegjøre for en helhetlig og formålsstyrt etterforskning som ivaretar krav til kvalitet, effektivitet, rettsikkerhet og forebygging av justisfeil – med utgangspunkt i en praktisk case de får utdelt på eksamensdagen. De har en time til individuelle forberedelser. Deretter blir de ført til en kommisjon bestående av tre sensorer som representerer hver sine fagelementer. Fagene jus og rapport- og etterforskningslære er representert i alle kommisjoner. I tillegg er ett av de tre øvrige fagelementer representert. Kandidatens prestasjon skal vurderes i et helhetlig perspektiv – også i lys av de fagelementer som ikke er representert.

Studenten får en halvtime til å presentere sin etterforskning, etterfulgt av et kvarter med spørsmål fra sensorene. Sensur gis på slutten av hver eksamensdag ved at kandidaten blir kalt inn til kommisjonen for å få sin karakter, og – dersom de ønsker – en begrunnelse for denne. Studentene kan ikke klage på karakteren, kun på formelle feil. Kommisjonen består av en blanding av eksterne fagfolk og interne fagansatte ved PHS. De interne er hentet fra bachelorutdanningen (1. og 3. trinn), samt fra PHS' avdeling for etter- og videreutdanning. Det varierer hvilket fagelement som er representert ved interne og eksterne sensorer, og kommisjonssammensetningen endres fra dag til dag. Noen kommisjoner består bare av interne sensorer, noen av en ekstern og to interne og (unntaksvis) to eksterne og en intern. Noen sensorer er med i hele perioden, mens andre kun deltar en dag eller to. Det hender man bytter sensorer på tvers av lokasjoner, men som oftest sensurerer man (også) der man har undervist.

¹ Siste studieår (B3)

Eksamen gjennomføres i løpet av en 2-ukers periode i juni. I Oslo er mellom 40 og 50 studenter (2 klasser) oppe til eksamen hver dag, fordelt på 10 kommisjoner. I tillegg gjennomføres tilsvarende eksamen ved PHS' lokasjoner i hhv. Bodø og Stavern i samme periode. Ved vårens gjennomføring var karaktersnittet i Oslo C, mens det i Bodø og Stavern var B.

I forkant av vårens eksamensgjennomføring ble det avholdt totalt fire sensormøter, hvorav to i Oslo, ett i Stavern og ett i Bodø. Hver sensor var forventet å delta på ett av møtene. Det er utarbeidet relativt omfattende vurderingskriterier² for etterforskningsemnet. Hver eksamensoppgave har fem sensorveiledninger, en for hvert fag. På sensormøtet jeg deltok på ble det påpekt at sensorene måtte sette seg inn i alle sensorveiledninger som var utarbeidet for de oppgavene de skulle sensurere, og særlig for de fagelementer som ikke var representert i kommisjonen. Vurderingskriterieskrivet ble ikke gjennomgått på møtet, men man gjennomgikk én sensorveiledning i fellesskap.

Jeg underviser selv innen etterforskningsemnet og har deltatt som sensor en rekke ganger. Min erfaring er at vurderinger gjøres ulikt – til dels svært ulikt – på tvers av kommisjoner og personer. Jeg ønsker jeg å få mer kunnskap om hvordan vurderingene foregår på tvers av sensorer og hva som vektlegges ved karaktersettingen. Vi har per i dag lite kunnskap om dette ved PHS, og undersøkelsen kan bidra til å gjøre ordningen mer transparent og å kaste lys over kvaliteten på vurderingsarbeidet ved PHS. Økt innsikt i hvordan ulike sensorer vurderer eksamensprestasjoner og hva som legges til grunn ved vurderingene, vil kunne skape godt beslutningsgrunnlag i det videre arbeidet ved PHS generelt og innen etterforskningsemnet spesielt. Regjeringen har klare forventninger til at utdanningsinstitusjonene sørger for at det er god sammenheng mellom det som skal læres og de vurderingsformer som benyttes (Meld. St. 16 (2016-2017), s. 24). Mitt prosjekt vil kunne gi viktig innsikt i dette arbeidet. Etterforskningsarbeid vil ofte innebære inngripen i enkeltpersoners skjebner, og feil kan få store konsekvenser både for enkeltindivider og rettssikkerheten i samfunnet. Politidistriktene må kunne ha tillit til at vurderingen gjenspeiler kandidatens kompetanse, slik at de kan gjøre riktige prioriteringer ved ansettelse. Kvaliteten på vurderingsarbeidet vil dermed kunne ha betydning for tilliten til både politiet og politiutdanningen. På individnivå vil karakteren i etterforskningsemnet kunne ha stor betydning for den enkelte student, fordi denne ofte vektlegges ved ansettelse av nyutdannede. Generelt bør det også være sammenheng mellom læringsinnsats og resultater. Hard jobbing for å oppnå forståelse og mestring – innenfor de fastsatte læringsutbytter – bør gi uttelling til eksamen. Studentenes rettssikkerhet er dermed også en viktig grunn til mitt ønske om å få dypere innsikt i kvaliteten på vurderingsarbeidet ved PHS.

Oversikt over forskning på feltet

Det synes å være lite forskning innen dette feltet i Norge, særlig knyttet til muntlig eksamen. Dobson skriver at det har vært gjort lite forskning både i Norge og internasjonalt for å klargjøre hva som foregår under muntlig eksamen, noe som delvis skyldes at det har vært vanskelig å få tilgang til systematisk observasjon av muntlige eksamenssituasjoner (Dobson, 2007b).

Bjølseth m.fl. har gjennomført et omfattende utviklingsarbeid ved en norsk profesjonsutdanning, med formål å øke kvaliteten i sensureringen. I sin artikkel viser de særlig til britisk forskning, og peker på at denne viser stor variasjon i sensureringsarbeidet og ofte dårlig sensorsamsvar (Bjølseth, Havnæs & Lauvås, 2011). I en norsk studie analyserte Raaheim sensureringen av essaybesvarelser i psykologi. Han fant at det var betydelig variasjon mellom sensorenes karaktersetting av samme besvarelser (Raaheim, 2000). Han har nylig utgitt bok om eksamen som vurderingsform, hvor han blant annet ser på forskningen som kartlegger sammenhengene mellom vurderingsform og læring (Raaheim, 2016).

² Skriv på 13 sider

Også Nederland har et sterkt forskningsmiljø innen vurderingsarbeid. Jeg har ikke rukket å sette meg inn i denne.

Jeg har ennå ikke lest veldig mye litteratur, men har gjort betydelige litteratursøk og fått tips fra blant andre Roger Bakken, som er i ferd med å slutføre sin doktoravhandling innen samme tema. Under listes opp noe forskning som kan være sentral for mitt prosjekt;

Stephen Dobson (2007a, 2007b, 2010, 2018)

Martin Johnson (2008)

Marianne Maugesten (2010, 2011)

Arild Raaheim (2000, 2016)

Elliot Eisner (2017)

Problemstilling og forskningsspørsmål

Etter å ha diskutert prosjektet med kolleger og læringsgruppen har jeg besluttet å avgrense til å se på interne sensorers vurderinger. Det er mange interne sensorer involvert ved gjennomføring av denne eksamen, og de har ulik bakgrunn og underviser ved ulike lokasjoner, trinn og avdelinger innenfor PHS. Mange av dem underviser innen etterforskningsemnet, men ikke nødvendigvis i 3. trinn. Jeg ønsker å få mer kunnskap om hvordan fagansatte innen egen utdanningsinstitusjon vurderer muntlig eksamen – med utgangspunkt i de samme prestasjoner. Gjøres vurderingene innenfor samme rammer, eller er det (for) store variasjoner og sprik i hva som vektlegges? Det er også en praktisk side i avgrensningen ved at det kan bli utfordrende å få eksterne sensorer til å sette av den tid som kreves for å delta i prosjektet. Hver sensor må sett av rundt 3 timer til deltakelsen.

På bakgrunn av ovennevnte har jeg valgt å undersøke problemstillingen;

Hvordan vurderer interne sensorer ved PHS muntlige eksamensprestasjoner (i emnet Etterforskning)?

Jeg ønsker å få mer kunnskap om hvilke vurderinger som gjøres av den enkelte, og hva som skjer i samspillet mellom sensorene når de skal enes om en felles karakter – slik man gjør ved gjennomføring av muntlig eksamen i etterforskningsemnet. Derfor har jeg valgt følgende to forskningsspørsmål;

1. Hvilke vurderinger legger de ulike vurdererne til grunn for sine forslag til karakter?
2. Hvordan forhandler sensorene seg frem til konsensus for karakter? (eller) *Hvordan vektlegges de ulike vurderingene i samspillet når man skal sette karakter?*

Teoretisk tilnærming

Jeg ønsker primært å gjøre en induktiv undersøkelse, hvor min empiri danner grunnlag for den teorien som skal inntas i oppgaven. Derfor har jeg per nå ikke klart for meg hvilken teori som skal inntas. På bakgrunn av det jeg har lest så langt har jeg imidlertid laget meg noen forslag til mulige teoretiske innfallsvinkler som kan vurderes etter hvert som prosjektet skrider frem. Disse er;

- A five dimensional framework for authentic assessment (Gulikens, Bastiaens & Kirschner, 2004). Så vidt jeg har forstått fokuseres det her på oppgavens autensitet og den sosiale og

fysiske settingen eksamen gjennomføres i. Hvilken innflytelse har det at det at vurderingen skjer i andre kontekster enn de autentiske? Tas det hensyn til at det er en kunstig situasjon?

- Om skjønn og skjønnsutøvelse i vurdering (Eisner, 2017)
- Kompetansebasert vurdering (Wolf, 1995)
- Helhetlig vurdering (Baartman, Bastiaens, Kirschner & Van der Vleuten, 2007; Bjølseth et al., 2011; Johnson, 2008; Vaughn, 1991)
- Kriteriebasert vurdering (Bjølseth et al., 2011)
- Aktivitetsteori (Johnson, 2008)

Metodisk tilnærming

Ved gjennomføring av etterforskningseksamen i juni 2018 fikk jeg samtykke av to studenter til å videofilme deres eksamensprestasjoner. Det er disse videoopptakene mine informanter vil bli bedt om å vurdere og karaktersette i den videre datainnsamlingen.

Metodevalget er gjort med utgangspunkt i hva som vil være mest hensiktsmessig for å belyse problemstillingen og tilhørende forskningsspørsmål. Jeg har valgt å kombinere bruk av spørreskjema og fokusgruppeintervju. Spørreskjema er valgt fordi det, innenfor mine rammer, gir meg mulighet til å hente data fra flere sensorer enn bruk av individuelle dybdeintervju ville gjort. Ved å bruke åpne spørsmål vil jeg også få belyst sensorenes begrunnelser, selv om jeg mister muligheten til å stille oppfølgingsspørsmål ved denne tilnærmingen. Derfor blir det svært viktig at jeg understreker viktigheten av å være detaljert i begrunnelsene til informantene før de fyller ut skjemaet. Dataene innhentet ved spørreskjema vil primært belyse forskningsspørsmål nr. 1, og vil bestå av spørsmål knyttet til karaktersetting og vurderingene som ligger til grunn for denne. I tillegg vil jeg inn ta noen spørsmål om sensor og dennes bakgrunn (blant annet arbeidssted og erfaring), slik at eventuelle mønstre mellom erfaring og/eller arbeidssted og vurderinger kan belyses.

For å belyse forskningsspørsmål nr. 2 vil jeg gjennomføre fokusgruppeintervju bestående av tre informanter som representerer ulike fagelementer. Tilnærmingen vil kunne bidra til å gjøre undersøkelsen så relevant som mulig sammenlignet med eksamensformen som er utgangspunkt for undersøkelsen, og for å få dypere innsikt i hva som skjer mellom personene i gruppen ved karakterfastsetting. Jeg vurderte også observasjon for dette formål. Imidlertid gir fokusgruppeintervju meg mulighet til å stille oppfølgingsspørsmål, noe som jeg anser som nødvendig for mitt formål. Det er ikke min observasjon av prosessene som det skal forskes på, men informantenes tanker og deres vurderinger og samspill. Derfor ønsker jeg å ha mulighet til å stille dem spørsmål om dette, eksempelvis om hva som gjorde at de ble enige som de gjorde. Jeg ser for meg at spørsmålene i fokusgruppeintervjuet primært vil komme til slutt, mens jeg innledningsvis setter rammene for diskusjonen hvor de skal diskutere seg frem til en felles karakter. Denne diskusjonen må ha en realistisk tidsramme i tråd med det som er tilfellet til eksamen i etterforskningsemne, kanskje rundt 10 minutter. Etter innspill fra læringsgruppen og veileder ser jeg for meg tre fokusgruppeintervju, og deretter en vurdering av behov for flere med fokus på metning. Fokusgruppeintervjuene vil bli gjennomført som videokonferanse, og bli tatt opp digitalt (lyd og bilde). Tilnærmingen er av både praktisk, ressursmessig og økonomisk art, idet sensorene er fordelt på lokasjonene Oslo, Bodø og Stavern (og en fra Kongsvinger). Ved valgte tilnærming kan jeg gjennomføre fokusgruppeintervju med sensorer som befinner seg på ulike steder i landet.

Fremdriftsplan

År	Periode	Gjøre hva	Status
2018	Før nyttår	Litteratursøk og litteraturlesning	Under arbeid
		Sende søknad til NSD	Under arbeid
		Lage informasjonsskriv til informanter	Under arbeid
		Fullføre spørreskjema og få tilbakemelding fra veileder på dette	Under arbeid
		Lage intervjuguide til fokusgruppeintervju	Under arbeid
		Skaffe informanter og gjøre avtaler (utvalg og henvendelse)	Ikke påbegynt
2019	Januar-februar	Lese litteratur og forskning	Under arbeid
		Datainnsamling (både spørreskjema og fokusgruppeintervjuer)	Ikke påbegynt
	Mars-april	Transkribere fokusgruppeintervjuer i NVivo	Ikke påbegynt
		Bryte ned og strukturere data fra spørreskjemaene i NVivo	Ikke påbegynt
	Mai-juli	Lese litteratur og forskning	Under arbeid
		Analysere dataene	Ikke påbegynt
		Litteraturlesing	Under arbeid
	2020	August-Mars	Skrive oppgaven
April		Litteraturlesing	Ikke påbegynt
		Ferdigstille oppgaven	Ikke påbegynt

Litteraturliste

- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A. & Van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks (s. 114-129).
- Bjølseth, G., Havnes, A. & Lauvås, P. (2011). Lavt sensorsamsvar – kan det bedres? *Uniped*, 34(04), 4-16.
- Dobson, S. (2007a). Teoretisering rundt muntlig eksamen - en kvalitativ tilnærming. *Norsk Pedagogisk Tidsskrift*, 91(02), 137-149.
- Dobson, S. (2007b). Teoretisering rundt muntlig eksamen – en kvalitativ tilnærming. *Norsk Pedagogisk Tidsskrift*, 91(02), 137-149.
- Dobson, S. (2010). Nasjonale og internasjonale utfordringer innen elev- og lærlingvurdering. I S. Dobson & R. Engh (Red.), *Vurdering for læring i fag* (s. 24-36). Kristiansand: Høyskoleforlaget.
- Dobson, S. (2018). *Assessing the viva in higher education : chasing moments of truth* (The enabling power of assessment, 1st edition 2018. utg., Bind 6): Springer International Publishing Springer.
- Eisner, E. W. (2017). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Teachers College Press.
- Gulikers, J., Bastiaens, T. & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86.
- Johnson, M. (2008). Exploring assessor consistency in a Health and Social Care qualification using a sociocultural perspective. *Journal of Vocational Education & Training*, 60(2), 173-187. doi: 10.1080/13636820802042446
- Maugesten, M. (2010). Muntlig eksamen-får studentene bedre vist hva de kan? En analyse av åtte studenters forståelse på muntlig eksamen i matematikk: rapport-NFR-prosjektet.
- Maugesten, M. (2011). Muntlig eksamen. En analyse av åtte studenters forståelse på muntlig eksamen i matematikk. *Norsk Pedagogisk Tidsskrift*(04), 260-272.
- Meld. St. 16 (2016-2017). (2017). *Kultur for kvalitet i høyere utdanning*. Hentet fra <https://www.regjeringen.no/contentassets/ae30e4b7d3241d5bd89db69fe38f7ba/no/pdfs/stm201620170016000dddpdfs.pdf>
- Raaheim, A. (2000). En studie av inter-bedømmer reliabilitet ved eksamen på psykologi grunnfag. *Tidsskrift for den norske psykologiforening*, 37(03), 203-213.
- Raaheim, A. (2016). *Eksamensrevolusjonen: Råd og tips om eksamen og alternative vurderingsformer*. Oslo: Gyldendal akademisk.
- Vaughn, C. (1991). Holistic assessment: What goes on in the rater's mind? I L. Hamp-Lyons (Red.), *Assessing second language writing in academic contexts* (s. 111-125). Norwood: NJ: Ablex.
- Wolf, A. (1995). *Competence-based assessment (Assessing assessment)*. Buckingham: Open University Press.

NSD NORSK SENTER FOR FORSKNINGSDATA

Meldeskjema 452212

Sist oppdatert

12.02.2019

Hvilke personopplysninger skal du behandle?

- Navn (også ved signatur/samtykke)
- Adresse eller telefonnummer
- E-postadresse, IP-adresse eller annen nettidentifikator
- Bilder eller videoopptak av personer
- Lydopptak av personer
- Bakgrunnsopplysninger som vil kunne identifisere en person

Type opplysninger

Du har svart ja til at du skal behandle bakgrunnsopplysninger, beskriv hvilke

Jeg vil stille noen spørsmål knyttet til sensors bakgrunn og erfaring. Disse vil være hvilket fag/emne(r) vedkommende underviser i, hvilken del av organisasjonen vedkommende tilhører og antall ganger vedkommende har sensurert i etterforskningsemnet.

Skal du behandle særlige kategorier personopplysninger eller personopplysninger om straffedommer eller lovovertridelser?

Nei

Prosjektinformasjon

Prosjekttittel

Vurderernes vurderinger: Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?

Prosjektbeskrivelse

Formålet er å få mer kunnskap om hvordan sensorer ved Politihøgskolen (PHS) vurderer muntlig eksamen i etterforskningsemnet i bachelorutdanningens 3. studieår, og hva som vektlegges ved karaktersetningen. Det er gjort lite forskning innen temaet 'Vurderernes vurderinger' i Norge, og vi har per i dag lite kunnskap om hvordan vi vurderer på tvers av sensorer ved PHS. Prosjektet kan bidra til å kaste lys over kvaliteten på vurderingsarbeidet ved PHS, og å gjøre ordningen mer transparent. Den kan også bidra til å skape godt beslutningsgrunnlag for det videre vurderingsarbeidet ved høgskolen generelt og innen etterforskningsemnet spesielt.

Problemstilling: Hvordan vurderer interne sensorer ved PHS muntlige eksamensprestasjoner (i emnet Etterforskning)?

Forsknings spørsmål:

1. Hvilke vurderinger legger de ulike vurdererne til grunn for sine forslag til karakter?
2. Hvordan forhandler sensorene seg frem til konsensus for karakter?

Fagfelt

Andre fagfelt

Begrunn behovet for å behandle personopplysningene

Viser til tidligere sendt meldeskjema med referanse 61040. I forlengelsen av 61040 vil jeg nå, som tidligere informert, innhente data fra sensorer innen etterforskningsemnet for å belyse problemstillingen. Dette vil gjøres gjennom spørreskjema og fokusgruppeintervju. I forbindelse med datainnsamlingen trenger jeg å kontakte informantene og å gjøre avtaler med dem. Derfor har jeg behov for å behandle personopplysningene navn, epostadresse og muligens også telefonnummer. Mine potensielle informanter befinner seg på ulike steder i landet (Oslo, Stavem, Bodø, Kongsvinger) og av ressursmessige årsaker (praktiske og økonomiske) jeg vil derfor gjennomføre fokusgruppeintervju som videokonferanse. Videokonferansene vil bli tatt opp på lyd. Fokusgruppeintervju som gjennomføres ved fysisk tilstedeværelse vil bli dokumentert med lyd.

Ekstern finansiering

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Christine Sætre Nordsletten, christinesatre@gmail.com, tlf: 41663269

Behandlingsansvar

Behandlingsansvarlig institusjon

OsloMet - storbyuniversitetet / Fakultet for lærerutdanning og internasjonale studier / Institutt for yrkesfaglærerutdanning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Hæge Nore, hnore@oslomet.no, tlf: 92442909

Skal behandlingsansvaret deles med andre institusjoner (felles behandlingsansvarlige)?

Nei

Utvalg 1

Beskriv utvalget

Sensorer i etterforskningsemnet

Rekruttering eller trekking av utvalget

Det er totalt ca 50 sensorer fordelt på lokasjonene Oslo, Bodø, Stavem og Kongsvinger som kan være aktuelle å spørre i henhold til min problemstilling og mine avgrensninger. Jeg ønsker å innhente data fra så

mange sensorer som mulig, og vil derfor sende ut mail til alle aktuelle med forespørsel om å delta i prosjektet. Blant dem som takker ja vil noen i tillegg bli spurt om å delta i fokusgruppeintervju (som skal finne sted like etter at videoen med eksamensbesvarelsen er sett og spørreskjemaet er utfylt). Jeg tenker som utgangspunkt rundt tre fokusgruppeintervju å tre sensorer. I disse intervjuene vil informantene bli bedt om å diskutere seg frem til/enes om en felles karakter for besvarelsen - tilsvarende slik det gjøres ved reell eksamensgjennomføring. De som deltar her må representere de tre fagelementer som var i den kommisjonen som kandidaten faktisk ble eksaminert av. I skriv om informert samtykke vil jeg gjøre rede for prosjektet, herunder at noen vil bli spurt om å delta på fokusgruppeintervju i tillegg til spørreskjemaet. Jeg vil også innta en avhuling på om hva de eventuelt ønsker å delta på; enten kun spørreskjema eller både spørreskjema og fokusgruppeintervju. For å velge kandidater til fokusgruppeintervju vil jeg trekke lodd blant de som ønsker å delta begge steder. Dersom det viser seg at veldig mange ønsker å delta på hele undersøkelsen og det viser seg å bli for omfattende for mine rammer, vil jeg også trekke lodd blant dem som har takket ja, for å velge informanter til den første delen (se video og fyll ut spørreskjema).

Jeg har huket av for videoopptak og lydopptak av personer under. Det er fordi jeg tenker å gjennomføre (noen av) fokusgruppeintervjuene via videokonferanse, og som jeg da vil ta opp digitalt (lyd og bilde). Grunnen til denne tilnærming er av både praktisk, ressursmessig og økonomisk art, idet sensorene er fordelt på lokasjonene Oslo, Bodø og Stavem (og en fra Kongsvinger). Ved valgte tilnærming kan jeg gjennomføre fokusgruppeintervju med sensorer som befinner seg på ulike steder i landet, og uten at det medfører store utgifter eller tidsbruk.

Alder

25 - 60

Inngår det voksne (18 år +) i utvalget som ikke kan samtykke selv?

Nei

Personopplysninger for utvalg 1

- Navn (også ved signatur/samtykke)
- Adresse eller telefonnummer
- E-postadresse, IP-adresse eller annen nettidentifikator
- Bilder eller videoopptak av personer
- Lydopptak av personer
- Bakgrunnsopplysninger som vil kunne identifisere en person

Hvordan samler du inn data fra utvalg 1

Elektronisk spørreskjema

Grunnlag for å behandle alminnelige kategorier av personopplysninger

Samtykke (art. 6 nr. 1 bokstav a)

Gruppeintervju

Grunnlag for å behandle alminnelige kategorier av personopplysninger

Samtykke (art. 6 nr. 1 bokstav a)

Informasjon for utvalg 1

Informerer du utvalget om behandlingen av opplysningene?

Ja

Hvordan?

Skriftlig informasjon (papir eller elektronisk)

Tredjepersoner

Skal du behandle personopplysninger om tredjepersoner?

Nei

Dokumentasjon

Hvordan dokumenteres samtykkene?

- Elektronisk (e-post, e-skjema, digital signatur)
- Manuelt (papir)

Hvordan kan samtykket trekkes tilbake?

Ved å kontakte meg. Epost er trolig det letteste for respondentene.

Hvordan kan de registrerte få innsyn, rettet eller slettet opplysninger om seg selv?

Ved henvendelse til meg.

Innsyn kan gis ved at vi treffes (fysisk eller via videokonferanse eller per telefon/epost) og jeg viser de opplysninger som er lagret om den enkelte. Dersom det bes om kopi vil jeg trekke ut det som gjelder for den registrerte. Ved ønsker om rettelser/sletting vil dette skje gjennom kommunikasjon mellom meg og den registrerte, og hvor ønskede endringer i etterkant fremvises til og godkjennes av den registrerte.

Totalt antall registrerte i prosjektet

1-99

Tillatelser

Skal du innhente følgende godkjenninger eller tillatelser for prosjektet?

Behandling

Hvor behandles opplysningene?

- Mobile enheter tilhørende behandlingsansvarlig institusjon
- Ekstern tjeneste eller nettverk (databehandler)

Hvem behandler/har tilgang til opplysningene?

- Student (studentprosjekt)

- Databehandler

Hvilken databehandler har tilgang til opplysningene?

Det er kun jeg som behandler dataene. De er lagret på passordbeskyttet plassering hvor kun jeg har tilgang på Politihøgskolens OneDrive for Business-konto. For å kunne gi respondentene (som befinner seg ved Politihøgskolens ulike lokasjoner i Norge), tilgang til videoopptak av eksamensprestasjon, blir disse lagt på mitt passordbeskyttede område på Politihøgskolens Mediasite. Herfra kan de gjøres tilgjengelig via link - når jeg bestemmer. Det er kun mottaker av link som kan se videoen, og kun i tidsrom definert av meg. Respondenten vil kun ha "lesetilgang" til videoen, og de vil verken kunne laste ned eller lagre denne, ei heller trykke på pause eller stopp. Sistnevnte fordi gjennomføringen skal være realistisk sammenlignet med slik muntlig eksamen forløper. Videoopptakene vil bli fjernet fra Mediasite når datainnsamlingen er ferdig. De blir kun gjort "viewable" i de korte tidspunkt når jeg har avtale med respondenter om at de skal se videoen, og deretter settes til "private" av meg, slik at linken ikke lenger fungerer. Link blir sendt respondenter idet de skal igang med vurderingen, ikke før. Respondenten bes i tillegg om å slette linken etter gjennomføring, selvom jeg som sagt i tillegg gjør den inaktiv. Jeg har fått hjelp av Politihøgskolens IT-avdeling til å sette alle innstillinger.

Tilgjengeliggjøres opplysningene utenfor EU/EOS til en tredjestat eller internasjonal organisasjon?

Nei

Sikkerhet

Oppbevares personopplysningene atskilt fra øvrige data (kodenøkkel)?

Ja

Hvilke tekniske og fysiske tiltak sikrer personopplysningene?

- Opplysningene anonymiseres
- Flerfaktoraутentisering

Varighet

Prosjektperiode

20.08.2018 - 20.06.2020

Skal data med personopplysninger oppbevares utover prosjektperioden?

Ja, data med personopplysninger oppbevares til: 20.06.2021

Til hvilket formål skal opplysningene oppbevares?

Forskning

Hvor oppbevares opplysningene?

Ved Politihøgskolen

Vil de registrerte kunne identifiseres (direkte eller indirekte) i oppgave/avhandling/øvrige publikasjoner fra prosjektet?

Nei

Tilleggsopplysninger

Utdypning av hvordan opplysningene skal behandles;

Filer med lydopptak av videokonferanse (fokusgruppeintervju) vil lagres på mitt passordbeskyttede område på Politihøgskolens OneDrive for Business-konto. Jeg har undersøkt med egen IT-avdeling og fått bekreftet at det kun er jeg som har tilgang her. Heller ikke ansatte ved IT-seksjonen har tilgang til mitt område. Navn og eventuell kontaktinformasjon vil bli anonymisert i alle transkriberte filer. Personopplysningene blir erstattet med en kode som også er knyttet til den enkelte respondent i koblingsnøkkelen. Fil med koblingsnøkkelen lagres på et adskilt og passordbeskyttet plassering hvor kun jeg har tilgang. Jeg har fått egen lisens til questback via Politihøgskolen, og vil bruke denne til spørreskjema. Det vil kun være meg som har tilgang til disse dataene.

Siden det trolig blir ganske mange informanter som skal gjøre en individuell vurdering av prestasjonen (ved å se video og besvare spørreskjema) blir det for omfattende for meg å møte dem fysisk og å ha med videoopptakene på den måten. Jeg vil derfor tilgjengeliggjøre videoopptaket via en link som kun gir "lese/visningstilgang" (og ikke muligheter for lagring eller nedlasting), og som kun er tilgjengelig i de 45 minuttene videoen spilles. Dette gjøres via Politihøgskolens Mediasite, hvor jeg har et eget passordbeskyttet område. Jeg har undersøkt med egen IT-avdeling og fått bekreftet at det kun er jeg som har tilgang her. Videoene vil bli fjernet fra Mediasite straks datainnsamlingen er ferdig.

Det er ønskelig å beholde dataene i ett år etter prosjektstutt slik at sensorer ved OsloMet kan ha mulighet til å se på disse dersom det skulle være ønskelig ifm sensur av masteroppgaven (vi ble anbefalt å be om dette på studiet - master i yrkespedagogikk, OsloMet).

Vurderernes vurderinger: Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?

Referanse

452212

Status

Vurdert

[Åpne Meldeskjema](#)
[Vurdering](#)

N

NSD Personvern

12.02.2019 15:45

Det innsendte meldeskjemaet med referansekode 452212 er nå vurdert av NSD.

Følgende vurdering er gitt:

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 12.02.2019, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD ENDRINGER

Dersom behandlingen av personopplysninger endrer seg, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringer gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 20.06.2021.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

I melding 05.02.2019 opplyser student om at det er innhentet samtykke til at videoopptak kan brukes i prosjektet, vår referanse 61040.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet

formålet med prosjektet.

- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTE RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Marianne Høgetveit Myhren
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

C

Christine Nordsletten

12.02.2019 12:46

Nå har jeg returnert oppdatert meldeskjema. Mvh Christine Nordsletten

N

NSD Personvern

12.02.2019 12:45

Kvittering på at meldeskjema med referansekode 452212 er innsendt og mottatt.

M

Marianne Høgetveit Myhren

12.02.2019 09:29

Hei igjen. Takk for tilbakemelding. Jeg har sjekket og alt er i orden. Fint om du returnerer meldeskjemaet til meg, slik at jeg kan ferdigvurdere prosjektet. Vennlig hilsen Marianne Myhren

C

Christine Nordsletten

05.02.2019 10:44

Hei, og takk for melding. Ja - både studenter og sensorer har fått informasjon om prosjektet, herunder formålet med opptaket og videre bruk av dette. De ble alle muntlig informert av meg, i tillegg til at de signerte skriv om signert samtykke utformet med utgangspunkt i mal på NSDs nettsider (i juni 2018). Disse er sendt NSD tidligere, i forbindelse med at prosjektet først ble meldt til NSD, se saksnummer 61040. Gi meg gjerne tilbakemelding dersom du ønsker

skrivene om informert samtykke (til henholdsvis studenter og sensorer på video) oversendt på nytt. Det er forøvrig kun studenten som vises på videoopptaket. Sensorene er "synlig" i form av lyd/stemme. Mvh Christine Sætre Nordsletten, tlf. 41663269.

N

NSD Personvern

04.02.2019 18:07

Det innsendte meldeskjemaet med referansekode 452212 må kompletteres for at NSD kan fortsette vurderingen.

Når du har gjort oppdateringene i skjemaet, må du gå til siden "send inn" og trykke "bekreft innsending".

Dersom du har ytterligere kommentarer eller spørsmål kan du skrive en melding i dialogfeltet over og trykke "send melding".

Følgende kommentar er gitt av NSDs personvernrådgiver:

Hei

Jeg har tatt en gjennomgang av meldeskjemaet ditt og det meste ser greit ut, men jeg har et spørsmål. Det fremgår at deltakerne skal se en videofilm reell muntlig eksamenspresentasjon. Jeg lurer da på om sensorer og studenter på filmopptakene er informert og evt. har gitt samtykke til at filmopptakene brukes på denne måten. Jeg ber om at du gir meg en tilbakemelding på dette.

Vennlig hilsen
Marianne H. Myhren

N

NSD Personvern

21.12.2018 16:30

Kvittering på at meldeskjema med referansekode 452212 er innsendt og mottatt.

v87

12.2.2019

Meldeskjema for behandling av personopplysninger

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Vurderernes vurderinger: Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?

Referansenummer

452212

Registrert

21.12.2018 av Christine Nordsletten - christinesatre@gmail.com

Behandlingsansvarlig institusjon

OsloMet - storbyuniversitetet / Fakultet for lærerutdanning og internasjonale studier / Institutt for yrkesfaglærerutdanning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Hæge Nore, hnore@oslomet.no, tlf: 92442909

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Christine Sætre Nordsletten, christinesatre@gmail.com, tlf: 41663269

Prosjektperiode

20.08.2018 - 20.06.2020

Status

12.02.2019 - Vurdert

Vurdering (1)

12.02.2019 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 12.02.2019, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD ENDRINGER

Dersom behandlingen av personopplysninger endrer seg, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringer gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 20.06.2021.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

I melding 05.02.2019 opplyser student om at det er innhentet samtykke til at videoopptak kan brukes i prosjektet, vår referanse 61040.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Marianne Høgetveit Myhren
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Vil du delta i forskningsprosjektet

”Vurderernes vurderinger: Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?»

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å få mer kunnskap om hvordan sensorer ved Politihøgskolen (PHS) vurderer muntlig eksamen i etterforskningsemnet i bachelorutdanningens 3. studieår, og hva som vektlegges ved karaktersettingen. Dette skrevet inneholder informasjon om målene for prosjektet og hva eventuell deltakelse vil innebære for deg.

Bakgrunn og formål

Undersøkelsen inngår i min masteroppgave ved masterstudium i yrkespedagogikk ved OsloMet. Det er gjort lite forskning innen temaet *Vurderernes vurderinger* i Norge, og vi har per i dag lite kunnskap om hvordan vi vurderer på tvers av sensorer ved PHS. Dette prosjektet kan bidra til å kaste lys over kvaliteten på vurderingsarbeidet ved PHS, og å gjøre ordningen mer transparent. Undersøkelsen kan bidra til å skape godt beslutningsgrunnlag for det videre vurderingsarbeidet ved høgskolen generelt og innen etterforskningsemnet spesielt.

I problemstillingen spør jeg «Hvordan vurderer interne sensorer ved PHS muntlige eksamensprestasjoner i emnet Etterforskning?». For å belyse problemstillingen ønsker jeg å se på 1) hvilke vurderinger de ulike sensorene legger til grunn for sine forslag til karakter og 2) hvordan sensorene forhandler seg frem til konsensus for karakter i kommisjonen.

Hvem er ansvarlig for forskningsprosjektet?

OsloMet er ansvarlig for prosjektet. Det er jeg som har designet undersøkelsen og vil stå for gjennomføringen i alle trinn. Førstelektor ved OsloMet, Hæge Nore, er min veileder.

Hvorfor får du spørsmål om å delta?

Du får spørsmål om å delta fordi du er i målgruppen for prosjektet. Med din erfaring og kompetanse kan du bidra til å kaste lys over problemstillingen. Jeg ønsker å samle data fra så mange interne sensorer som mulig. Derfor henvender jeg meg til alle aktuelle deltakere.

Hva innebærer deltakelse for deg?

Deltakelse innebærer å se en videofilmet reell, muntlig eksamensprestasjon som varer i rundt 45 minutter. På bakgrunn av filmen skal du besvare et spørreskjema med spørsmål om eksamensprestasjonen. Her blir du bedt om å karakter sette prestasjonen og begrunne karakteren. Spørreskjemaet inneholder også spørsmål om din erfaring, herunder hvilket fag/emne(r) du underviser innen, hvilken del av organisasjonen du tilhører og hvorvidt du har vært sensor i etterforskningsemnet i B3 tidligere. Utfylling av spørreskjemaet vil ta inntil 30 minutter, og vil bli gjennomført umiddelbart etter at du har sett videoen. Spørreskjemaet er elektronisk. Dine svar lagres digitalt. Enkelte deltakere vil bli invitert til å delta i fokusgruppeintervju. Der skal deltakeren diskutere seg frem til en felles karakter med to andre sensorer. Fokusgruppen vil bestå av personer som representerer hvert av de fagelementer som var representert i kommisjonen som vurderte den videofilmede eksamensprestasjonen. Fokusgruppeintervjuet vil ta cirka 15-30 minutter, og vil gjennomføres like etter at spørreskjemaet er utfyllt. Det vil gjøres enten ved fysisk møte eller via videokonferanse, og vil bli dokumentert på lyd (fysisk møte) eller video (videokonferanse).

Prosjektet er forankret hos ledelsen ved PHS. Det vil bli gitt tid til å delta i prosjektet.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake, uten å oppgi noen grunn. Dine data vil da bli slettet fra prosjektet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Dersom du deltar i fokusgruppeintervju (sammen med to andre) og i ettertid velger å trekke din deltakelse, vil dine svar bli fjernet fra det transkriberte datagrunnlaget, og ikke bli brukt videre. Selve opptaket av intervjuet vil imidlertid bli beholdt til prosjektets slutt, siden det også vil inneholde data som inngår i prosjektet i form av det som er sagt av de øvrige deltakerne i fokusgruppeintervjuet.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrevet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Alle filer (svar på spørreskjema, opptak av fokusgruppeintervju og transkriberte filer) vil bli lagret på passordbeskyttet plassering hvor kun jeg har tilgang. Dine personopplysninger vil lagres adskilt fra øvrige data. Navn og kontaktopplysninger vil bli erstattet med en kode som lagres på en og adskilt liste. I utgangspunktet er det ingen andre enn meg som vil ha tilgang til dataene, men OsloMet kan be om å få tilgang til de transkriberte filene i forbindelse med sensur av masteroppgaven. I så tilfelle vil disse bli delt med aktuell(e) person(er) der, men ditt navn og kontaktopplysninger vil være anonymisert i disse filene.

Du vil ikke bli omhandlet med navn eller andre personopplysninger i masteroppgaven, og vil ikke kunne gjenkjennes.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes medio juni 2020. Deretter skal masteroppgaven sensureres. Etter at endelig eksamensresultat foreligger, vil alle dine personopplysninger bli anonymisert. Lyd/videoopptak, spørreskjema og alle transkripsjoner slettes da. Etter anbefaling fra OsloMet vil jeg be NSD/lokalt personvernombud ved OsloMet om tillatelse til å lagre opplysningene til medio juni 2021, slik at OsloMet kan etterprøve eller gjennomgå dataene dersom det skulle være ønskelig eller nødvendig i forbindelse med sensur av masteroppgaven.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra OsloMet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Dersom du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Christine Sætre Nordsletten (undertegnede) eller prosjektansvarlig ved OsloMet: Hæge Nore

- OsloMet vil i nær fremtid ansette et lokalt personvernombud, men denne funksjonen er per dags dato ikke i aktiv funksjon.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Prosjektansvarlig
(Forsker/veileder)

Christine Sætre Nordsletten
(Student)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Vurderernes vurderinger: Hvordan vurderes muntlige eksamensprestasjoner ved Politihøgskolen?*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- Å delta ved å se video og besvare spørsmål i spørreskjema
- Å delta ved å se video, besvare spørreskjema og delta i fokusgruppeintervju (alle tre deler skjer rett etter hverandre)
- At mine personopplysninger lagres inntil et år etter prosjektslutt (til juni 2021), slik at datagrunnlaget for prosjektet kan deles med sensorer ved OsloMet dersom det bes om det i forbindelse med sensur av masteroppgaven. I de data som eventuelt deles vil mine personopplysninger være anonymisert, og jeg vil ikke kunne gjenkjennes.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 1. juni 2021.

(Signert av prosjektdeltaker, dato)

Vurdering - Del 1

I denne delen skal du kun besvare ett spørsmål. Deretter skal du se resten av videoen og besvare noen flere spørsmål. Siden jeg ikke har anledning til stille deg oppfølgingsspørsmål, er det viktig at du er konkret, detaljert og utfyllende i dine svar og begrunnelser.

Spørsmål om vurderingen

1) * Du har nettopp sett kandidatens frie redegjørelse (de første 30 minuttene). Kommisjonen har nå 15 minutter til å stille spørsmål til kandidaten, hvorav ca. 5 minutter er forbeholdt deg. Hvilke(t) spørsmål ville du ha stilt til kandidaten, og hvorfor?

0/4000

Send

100 % fullført

© Copyright www.questback.com. All Rights Reserved.

Vurdering - Del 2

Du skal nå besvare spørsmål om din vurdering av eksamensprestasjonen du nettopp har sett (30 + 15 minutter). Siden jeg ikke har anledning til stille deg oppfølgingsspørsmål, er det viktig at du er konkret, detaljert og utfyllende i dine svar og begrunnelser.

NB! Det er ikke mulig å navigere tilbake til en tidligere side, og du må derfor gjøre deg helt ferdig på en side før du velger "Neste".



Spørsmål om vurderingen forts.

1) * Hvilken karakter ville du gitt kandidaten? Her må du legge til grunn hele den prestasjonen du nå har sett, selvom spørsmålene som ble stilt kandidaten kanskje er andre enn dem du ville ha stilt

A B C D E F

2) * Beskriv detaljert de vurderinger og begrunnelser du legger til grunn for karakteren du har gitt

0/4000



Spørsmål om vurderingen forts.

3) Har du lest pensum for det fagelement du representerer i kommisjonen? NB! Flere avkryssninger er mulig. Bruk "Annet"-alternativet til å utdype svaret hvis behov.

Ja Nei Delvis Annet

4) Har du lest pensum for øvrige fag i etterforskningsemnet i B3 enn det du selv representerer i kommisjonen? NB! Flere avkryssninger er mulig. Bruk "Annet"-alternativet til å utdype svaret hvis behov.

Ja Nei Delvis Annet


5) * Det er utarbeidet fem sensorveiledninger til denne oppgaven - en per fagelement. Hvordan brukte du sensorveiledningene ved din vurdering av eksamensprestasjonen?

0/4000

6) * Hvordan brukte du vurderingskriteriene som er utarbeidet for etterforskningsemnet ved din vurdering av eksamensprestasjonen?

0/4000

7) * Hvordan brukte du beskrivelsene av bokstavkarakterene - slik de står beskrevet i Forskrift om studier og eksamen ved PHS - ved din vurdering?



0/4000



Spørsmål om sensor/erfaring

8) * Hvilken PHS-lokasjon tilhører du?

- Oslo
- Bodø
- Stavem
- Kongsvinger

9) * Hvilken del av organisasjonen er du tilknyttet? NB! Flere avhukinger er mulig. Bruk "Annet"-alternativet til å utdype svaret hvis behov.

- B1
- B2
- B3
- EVU
- Annet

10) * Underviser du innen emnet etterforskning? NB! Flere avhukinger er mulig. Bruk "Annet"-alternativet til å utdype svaret hvis behov.

- Ja, i B1
- Ja, i B3
- Ja, på EVU
- Nei
- Annet

11) * Hvilket fagelement representerer du i kommisjonen?

- Jus
- Rel
- Krimtek
- Psy
- Dipo

12) * Har du vært sensor i etterforskningsemnet i B3 tidligere? Eventuelt (ca) hvor mange ganger?

13) Har du kjennskap til denne kandidaten, for eksempel gjennom undervisning eller veiledning? NB! Flere avhukinger er mulig. Bruk

7.9.2019

QuestBack

"Annet"-alternativet til å utdype svaret hvis behov.

Ja Nei Annet

Send

100 % fullført

© Copyright www.questback.com. All Rights Reserved.

Hvordan vurderer interne sensorer ved PHS muntlige eksamensprestasjoner (i emnet Etterforskning)?			
Forskningsspørsmål	Innledende informasjon	Intervju spørsmål	Egne merknader
Hvordan vektlegges de ulike vurderingene i samspillet når man skal sette karakter?	Dere tre er nå å anse som en eksamenskommisjon som skal vurdere og karaktersette den prestasjonen dere nettopp har sett og individuelt vurdert. Jeg ønsker at dere gjør som dere pleier, eller ville gjort, som kommisjon ved en reell eksamen etter at kandidaten har forlatt rommet. Det er fint om dere holder dere innenfor en tidsramme på et kvarter, fordi det er realistisk sammenlignet med reell eksamen. Til slutt, etter at dere har satt karakter som kommisjon, har jeg kanskje et eller noen få spørsmål. Før det kommer jeg ikke til å si noe som helst. Jeg har ikke plukket ut en kommisjonsleder, så dere får alle være ansvarlige for tiden. Da overlater jeg ordet til dere.	(Etter at kommisjonen har blitt enige om en karakter, og avhengig av innholdet i de samtaler/diskusjoner som har funnet sted); Hva var grunnen til at noen gikk med på å gå opp/ned i karakter sammenlignet med den individuelle karakteren de mente var riktig? Hva var grunnen til at noe ble vektlagt mer enn noe annet?	Oppfølgingsspørsmålene må vurderes i lys av det som har fremkommet under de samtaler/diskusjoner som "kommisjonen" har hatt for å enes om en felles karakter. Dersom det er stor enighet om karakteren kan det være sentralt å belyse hva som dannet grunnlag for enigheten. Ved uenighet er det sentralt å få mer informasjon om hva uenighetene består i og hva som gjorde at man vektla noe mer enn noe annet.

Christine Sætre Nordsletten

Fra: Christine Sætre Nordsletten
Sendt: torsdag 17. januar 2019 12:55
Emne: Deltakelse i forskningsprosjekt_vurdering av muntlig eksamen
Vedlegg: 08_Informasjonskriv_deltakelse i forskningsprosjekt.pdf

Hei.

Jeg jobber som REL-lærer i B3 Oslo og er også student ved OsloMets masterstudium i yrkespedagogikk. Jeg er nå i gang med min masteroppgave, hvor jeg skal undersøke hvordan muntlige eksamensprestasjoner vurderes ved PHS. Vi har per i dag lite kunnskap om hvordan vi vurderer på tvers av sensorer ved PHS, og prosjektet kan bidra til å få dypere innsikt i dette og å skape godt beslutningsgrunnlag for veien videre. Jeg vil undersøke problemstillingen med utgangspunkt i etterforskningsemnet, og fikk ved siste eksamensgjennomføring samtykke fra to studenter til å videofilme deres eksamensprestasjon. Disse vil danne utgangspunkt for min datainnsamling.

Du mottar denne mailen fordi du er/har vært sensor i etterforskningsemnet og dermed kan bidra til belyse problemstillingen.

Se vedlagte skriv for mer informasjon om prosjektet. Deltakelse innebærer at du må sette av noen timer en dag. Du kan derfor delta fra den lokasjonen du tilhører, og trenger ikke å reise. Dataene som samles fra den enkelte sensor vil inngå i et større datagrunnlag. Ingen vil bli omhandlet med navn eller andre personopplysninger i oppgaven, og ingen vil på noen måte kunne gjenkjennes.

Jeg setter stor pris på snarlig tilbakemelding på om du ønsker å delta.

Mvh
Christine Sætre Nordsletten
Politioverbetjent

Politi
Bacheloravdelingen

Telefon: (+47) 23 19 99 44
Mobil: (+47) 41 66 32 69
E-post: chrino@phs.no
www.phs.no

 please consider the environment before printing this e-mail



Christine Sætre Nordsletten

Fra: Christine Sætre Nordsletten
Sendt: tirsdag 29. januar 2019 14:25
Emne: Vurdering av eksamensprestasjon - mandag 4.februar
Vedlegg: SENSORMAPPE 23K.pdf

Hei. Tusen takk for at dere har sagt dere villig til å delta i forskningsprosjektet om vurdering av muntlig eksamen.

På mandag (04.02) skal dere se på video av en reell eksamensgjennomføring. Dette delen av prosessen gjøres individuelt.

Kandidaten dere skal se kommer opp i oppgave 23 K. Gjennomføringen skal være så realistisk som mulig, og jeg vedlegger derfor sensormappe med alle de dokumenter som ble brukt ved vårens gjennomføring. For enkelthets skyld har jeg fjernet dokumenter som er knyttet til andre oppgaver enn 23K (dvs oppgavetekster og sensorveiledninger). Det er ønskelig at dere forbereder dere som ved en reell gjennomføring.

Vi møtes på mitt kontor kl 9 for kort info. Deretter ser dere videoen/eksamensprestasjonen og besvarer spørsmål i (digitalt) spørreskjema individuelt. Link til video og spørreskjema får dere av meg mandag morgen.

Etter den individuelle vurderingen samles vi til et kort fokusgruppeintervju. Dere er da de tre sensorene utgjør denne eksamenskommissjonen (bestående av rel, jus og krimtek). I fokusgruppeintervjuet skal dere enes om en felles karakter for eksamensprestasjonen, som dere også ville gjort ved en reell gjennomføring. Intervjuet tar ikke lang tid.

Kan dere også signere på samtykkeskjemaet som var vedlagt forrige mail, samt huke av på de relevante «boksene»? Skjemaet kan enten scannes og sendes meg per mail, eller tas med på mandag.

Jeg er på mastersamling resten av uken, men send meg gjerne en mail eller ring dersom dere lurer på noe ☺

Ønsker dere en flott uke! 😊

Mvh
Christine Sætre Nordsletten
Politioverbetjent

Politi høgskolen
Bacheloravdelingen

Telefon: (+47) 23 19 99 44
Mobil: (+47) 41 66 32 69
E-post: chrino@pfs.no
www.pfs.no

 please consider the environment before printing this e-mail



Nivå 1	Nivå 2	Kode definisjon
Nedbryting trinn 3		Alle data brutt ned og kategorisert i runde 3 - teorinært i den forstand at kategoriene er laget med utgangspunkt i teorien. Teorien har fremkommet gjennom arbeidet med trinn 1 og 2
Kriterier og standarder		Overordnet kategori. Uttalelser som sier noe om hvilke "rammer" informantene legger til grunn for bedømmingen. Har tatt med flere av kategoriene som ble opprettet i nedbrytingens trinn 1 og 2, som "Gått med saken", "Undervisning" (hvorvidt casetyper/straffebudet er dekket i undervisningen), Vurderingskriterier, Pensum, Sensorveiledninger, Karakterbeskrivelser, andre..
Kriterier og standarder	Vurderingskriterier	Alle svar på spørsmål om bruk av vurderingskriterier og uttalelser der vurderingskriteriene er omhandlet i karakterbegrunnelsene
Kriterier og standarder	Sensorveiledninger	Alle svar på spørsmål om bruk av sensorveiledninger og uttalelser der sensorveiledningene er omhandlet i karakterbegrunnelsene
Kriterier og standarder	Pensum	Alle svar på spørsmål om bruk av pensum og uttalelser der pensum er omhandlet i karakterbegrunnelsene
Kriterier og standarder	Karakterbeskrivelser	Alle svar på spørsmål om bruk av karakterbeskrivelser og uttalelser der karakterbeskrivelsene er omhandlet i karakterbegrunnelsene
Kriterier og standarder	Gått med saken	Brukes der sensor vurderer på bakgrunn av spørsmålet hvordan det ville ha gått med saken
Vurderingsresultat (begrunnelser og argumentasjon)		Overordnet kategori. Karakteristika ved begrunnelsene og tilhørende argumentasjon. Fokus både på hvordan det argumenteres og særtrekk ved begrunnelsene. Flere av kategoriene inneholdende mulige særtrekk fra nedbrytingen trinn 1 og 2, for eksempel fag vs helhetlig fokus, konkrete vs generelle, er tatt med videre som underkategorier til de mer teoretisk orienterte kategoriene. Kategorien favner også uttalelser som belyser hvordan det argumenteres, både individuelt og i samspill med andre, og sammenhengen (evt mangel på sådan) mellom argumenter og slutninger.
Vurderingsresultat (begrunnelser og argumentasjon)	Etos	Uttalelser som benytter seg av appellformen Etos (talerens karakter og egenskaper ved taleren)
Vurderingsresultat (begrunnelser og argumentasjon)	Patos	Uttalelser som bruker følelsesuttrykk i argumentasjonen
Vurderingsresultat (begrunnelser og argumentasjon)	Logos	Uttalelser som benytter seg av appellformen Logos (logisk argumentasjon med sammenheng mellom argumenter og slutning)
Vurderingsresultat (begrunnelser og argumentasjon)	Spørsmål	Favner alle uttalelser som sier noe om spørsmålsstillingen og/eller betydningen av denne for vurderingen.
Vurderingsresultat (begrunnelser og argumentasjon)	Hel vs del	Sammenslåing av de to tidligere nodene "Helhet" og "Faginndelt". Favner uttalelser som belyser om vurderingen er helhetlig eller ikke. Altså helhetlig eller faginndelt.
Vurderingsresultat (begrunnelser og argumentasjon)	Struktur	Uttalelser som omhandler struktur-elementer som en del av vurderingen, altså forhold som går på ikke det rent faglige, men formen på fremføringen
Sitater		Uttalelser som det kan være aktuelt å bruke for å illustrere funn
Subjektivitet		Denne kategorien er opprettet til slutt som et resultat av funnene om at vi vurderer ulikt, og er ikke brukt som egen kategori under nedbrytingen