

# HELVYRD

## Helhetlig vurdering i et yrkesdidaktisk perspektiv

En studie av vurderernes helhetlige vurderinger av siste års lærlinger i elektrikerfaget i ulike vurderingssystemer innen fag- og yrkesopplæringen

**Roger Bakken**

OsloMet Avhandling 2020 nr 23

**OSLO METROPOLITAN UNIVERSITY**  
STORBYUNIVERSITETET





# HELVYRD

## Helhetlig vurdering i et yrkesdidaktisk perspektiv

En studie av vurderernes helhetlige vurderinger av siste års lærlinger i elektrikerfaget i ulike vurderingssystemer innen fag- og yrkesopplæringen

Roger Bakken

**OSLOMET**

Avhandling Ph.D. utdanningsvitenskap for lærerutdanning  
Institutt for yrkesfaglærerutdanning (YLU)  
Fakultet for lærerutdanning og internasjonale studier  
OsloMet – storbyuniversitetet

Høsten 2020

CC-BY-SA versjon 4.0

OsloMet Avhandling 2020 nr 23

ISSN 2535-471X(trykt)

ISSN 2535-5414 (online)

ISBN 978-82-8364-263-6 (trykt)

ISBN 978-82-8364-288-9 (online)

OsloMet – storbyuniversitetet

Universitetsbiblioteket

Skriftserien

St. Olavs plass 4,

0130 Oslo,

Telefon (47) 64 84 90 00

Postadresse:

Postboks 4, St. Olavs plass

0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

## Forord

I gjennomføring av en avhandling er man avhengig av flere personer som bidrar på ulike måter. Disse må takkes!

Først og fremst må jeg takke mine veiledere Professor Leif Christian Lahn og Professor Ronny Sannerud for tett og god veiledning. Dere har utfyllt hverandre på en god måte, noe som har vært vesentlig for å få inn ulike synspunkter og perspektiver på arbeidet underveis.

Videre vil jeg takke de øvrige deltakerne i MECVET prosjektet. Dosent Hæge Nore har vært med i hele prosjektperioden og har underveis gitt svært nyttige og konstruktive innspill til mitt arbeid. Førsteamanuensis Hedvig Skonhøft Johannesen har også i periodevis bidratt med gode og nyttige innspill. Andre prosjektdeltakere som har bidratt innledningsvis i mitt arbeide er Dr. Lars Heinemann, Universitetet i Bremen og Professor Trine Deichman-Sørensen, OsloMet.

Videre vil jeg takke mine informanter fra de tre ulike vurderingssystemene som har stilt opp. Uten dem hadde ikke avhandlingen blitt til!

En stor takk til ledelsen ved Institutt for Yrkesfaglærerutdanning som har lagt forholdene godt til rette for mitt arbeid. Underveis har jeg måtte forholde meg til flere ledere, og alle har vært positive og støttende slik at jeg har fått fullført arbeidet.

Universitetsbiblioteket ved OsloMet, avdeling Kjeller har bidratt svært bra. Spesielt vil jeg rette en stor takk til Universitetsbibliotekar Lilja Johannessen. I tillegg har Førsteamanuensis Kristin Solli bidratt med viktige innspill i innspurten av arbeidet.

Jeg vil rette en takk til mine kollegaer ved Yrkesfaglærerutdanningen på Teknologiske fag. Dere har vært en god støtte for at min undervisningsplikt kunne gjennomføres på en god måte. En spesiell takk til Universitetslektor Kurt Stokke som i perioder har lånt ut sin «lille røde stuga» i Sverige, hvor jeg har fått jobbet godt i mitt «reflektorium».

Min familie vil alltid ha betydning og intet unntak denne gangen. Tusen takk til dere!

Mine venner har spilt en viktig rolle for meg, selv om de kanskje ikke er like klar over det selv. Kontakten har vært begrenset i perioder, men vi får ta igjen dette med nye opplevelser i fremtiden. Takk til dere!



## Sammendrag

Min avhandling, *Helhetlig vurdering i et yrkesdidaktiske perspektiv* (HELVYRD), er en studie av vurderernes helhetlige vurderinger av lærlinger innenfor fag- og yrkesopplæringen i Norge. I dette systemet brukes ulike vurderingsformer gjennom hele utdanningsløpet hvor fagprøven etter 4 års løp (to år skole pluss to år læretid, 2+2-modell (normalt, men med avvik)) er en avsluttende vurdering. HELVYRD er en del av det NFR (Norsk Forskningsråd)-støttede MECVET-prosjektet (Measuring Competence in Vocational Education and Training), som har utprøvd en storskalavurdering med den tyske KOMET-modellen (Kompetenzdiagnostik im Berufsfeld Elektrotechnik) tilpasset den norske fag- og yrkesopplæringen.

I HELVYRD forskes det på vurderernes helhetlige vurdering i elektrikerfaget hvor tre vurderingssystemer anvendes, MECVET-modellen og to norske systemer: sentralt gitt skriftlig eksamen og fagprøve. Studien søker å finne ut av hvordan vurderere innenfor ulike vurderingssystemer kan vurdere en helhetlig og kompleks yrkeskompetanse innenfor elektrikerfaget. Ut i fra det har jeg utviklet en definisjon på helhetlig vurdering:

Helhetlig vurdering er å trekke inn ulike aspekter av en handling knyttet til en yrkesutøvelse i en konkret kontekst, der vurderer bruker sin erfaring, kompetanse og skjønn for i størst mulig grad å vurdere yrkeskompetansen opp mot de krav som stilles for utøvelsen av et yrke.

Studiens tre forskningsspørsmål lyder:

*-Hvordan kan vi forstå helhetlig vurdering i et yrkesdidaktisk perspektiv?*

Spørsmålet retter søkelyset mot et tema som i avhandlingen blir gjenstand for teoretisk drøfting og som gir en begrepsramme (*konseptuelt*) for å analysere det empiriske materialet.

*-I hvilken grad foretar vurdererne helhetlige vurderinger i de utvalgte vurderingssystemene?*

Dette spørsmålet skal besvares gjennom å beskrive innslaget av helhetlig vurdering (*deskriptivt*) i de tre vurderingssystemene.

*-Hvilke faktorer påvirker vurderernes helhetlige vurdering av siste års lærlingbesvarelser?*

Hvordan påvirker vurderingssystemene på vurderernes helhetlige vurdering (*forklarende*)?

Av de ulike faktorene som skal undersøkes, vil vurderingssystemene stå sentralt og dermed også de yrkesdidaktiske prinsippene som disse systemene bygger på.

Det teoretiske rammeverket er delt i to hovedområder: yrkesdidaktikk og helhetlig vurdering. Den yrkesdidaktiske diskusjonen tar utgangspunkt i både den internasjonale (nordisk og tysk) og nasjonale. Dette utvikles inn i seks perspektiver om yrkesdidaktikk basert på Gessler og Herrera (2015) sine seks forutsetninger. Teoretisk rammeverk om helhetlig vurdering har ikke vært mulig å bygge på et spesielt teoretisk utgangspunkt, men fokus har vært på kompetansebasert vurdering (Wolf, 1995) og annen tilstøtende teori. Den utvalgte litteraturen har blitt systematisert med støtte i den didaktiske relasjonsmodellen og utgjør ulike aspekter ved helhetlig vurdering.

Studiens forskningsdesign og metode er utviklet rundt kvalitative intervjuer som ble gjennomført med 13 vurderere fordelt på tre vurderingssystemer med støtte i tenkningen om «stimulated-recall» (Calderhead, 1981; Lyle, 2003) og «think-aloud» (Patton, 2015; Kuusela og Paul, 2000). Dokumentanalyse utgjør en mindre del av datainnsamlingen. Analyse av empiri skjer ved koding og kategorisering og har skjedd med støtte i tenkningen til Charmaz (2014) og Strauss og Corbin (1998). Presentasjon og diskusjon av resultatene har skjedd i de fremkomne temaer og kategorier. En ytterligere sammenfattende diskusjon gjøres i fire hoveddimensjoner.

Presentasjonen av studiens funn tilknyttet de tre forskningsspørsmålene:

- I studiens funn omkring forståelsen av helhetlig vurdering i et yrkesdidaktisk perspektiv (konseptuelt) kom det frem at lite har blitt skrevet om tematikken. Gjennom min utvikling av et teoretisk rammeverk om helhetlig vurdering fremkom ulike perspektiver, noe som ble ytterligere belyst gjennom det yrkesdidaktiske rammeverket. Det handler om å belyse en variant av vurdering for fag- og yrkesopplæring, hvor vurdereren ser ulike aspekter av en kompleks yrkesutøvelse i en bestemt kontekst gjennom å anvende sin erfaring, kompetanse og skjønn og vurdere opp imot de krav som stilles for utøvelsen. Helhetlig vurdering vil romme et syn på vurdering som handler om å fremme den faglige helheten.

Gjennom å belyse helhetlig vurderingen i utvalgte yrkesdidaktiske perspektiver kommer det til syne muligheter, utfordringer og begrensninger. Spesielt den yrkesdidaktiske tilhørigheten og påvirkningen fra yrkeslivet vil være sentralt. Yrkenes stadige utvikling kommer tydelig frem, noe som preger både vurderingsfeltet og det yrkesdidaktiske feltet isolert sett, men også hvordan disse kan ses i sammenheng. Det igjen preger forståelsen av hva som ligger i helhetlig vurdering.



- Studiens funn viser at vurderernes grad av helhetlig vurdering var høy. Det kan forklares ut fra vurderernes tilnærming til elektrikerfagets innholdselementer, anvendelsen av egen faglig bakgrunn og hvordan de anvendte vurderingskriterier inn i vurderingskonteksten. Det som fremmer den helhetlige vurderingen, er evnen til å anvende elektrikerfaget og yrkeslivet inn i vurderingen av besvarelsene. Funn viser at det å være en del av et vurderingsfellesskap er sentralt, og her blir den enkelte vurderer bevisstgjort og kan utvikle seg. Det er viktig for å holde seg oppdatert innenfor et elektrikerfag som er i stadig utvikling.

- Studiens funn om påvirkning av vurderernes helhetlige vurderinger gjelder først og fremst vurderingssystemenes former og elektrikerfagets utvikling. Vurderingssystemene påvirker vurdererne gjennom oppgavens utforminger, strukturer for vurderingsprosessene og de rammene som gis. Ulikheten mellom systemene er store, og dermed påvirkes vurdererne i ulik grad. Funn viser at jo strammere og mindre fleksibelt systemet er, jo mer innsnevrer det vurderernes mulighet til å operere helhetlig i sine vurderinger. Vurderernes medvirkning og eierforhold til vurderingsprosessen reduseres og medfører dermed i større grad en uniformering og mekanisering. Vurdererne vil utelate vesentlige sider ved det å vurdere en helhetlig yrkeskompetanse. Vurderingssystemene må tilpasse seg elektrikerfagets raske utvikling og gi vurdererne handlingsrom til å vurdere aktuell yrkeskompetanse. Sentrale funn viser at anvendelse av casebaserte oppgaver med form av en arbeidsordre og det å kunne delta i et vurderingsfellesskap med mulighet for å medvirke i utvikling og anvendelse av vurderingskriterier, var noe som påvirket positivt for i større grad å kunne gjøre helhetlige vurderinger.

Avslutningsvis viser studien at det vil være behov for ytterligere forskning på flere områder. Det omfatter forskning på vurdering i flere fag, flere former for vurderingssystemer, utvikling og anvendelse av mål og kriterier, betydningen av skriftlighet og ikke minst utviklingen av faglig innhold.

## Summary

My thesis, *Holistic Assessment in a Vocational Didactic Perspective (HELVYRD)*, is a study of assessors' holistic assessment of apprentices in vocational education and training (VET) in Norway. Different forms of assessment are used throughout the education and training, where the final trade examination, after four years (two years school-based and two years work-based apprentice, normal, but with some deviation) is a summative assessment. HELVYRD is a part of the MECVET-project (Measuring Competence in Vocational Education and Training), supported by the NFR (Norwegian Research Council), which has tried out large-scale assessment using the German COMET-model (Kompetenzdiagnostik im Berufsfeld Elektrotechnik) adapted to the Norwegian context.

The HELVYRD study is conducting research on assessors' holistic assessment in electrician where three different assessment systems are used, the MECVET-model and two Norwegian systems: centrally given written examinations and trade examinations. The study seeks to find out how the assessors within the three different assessment systems can assess the holistic and complex vocational competence of a skilled electrician. In my theoretical review, I have developed my own definition of holistic assessment:

Holistic assessment is to incorporate various aspects of work activity related to a professional practice in a specific context, where the assessor uses his/her experience, expertise and judgment to assess, as much as possible, the professional competence against the requirements for the profession.

The three research questions in this study are:

***How can we understand holistic assessment in the perspective of vocational didactics?***

This question focuses on an area which, in the thesis, is the subject of theoretical discussion and which provides a conceptual framework for analysing the empirical material.

***To what degree do the assessors conduct holistic assessments in the selected assessments systems?***

This question is answered by describing the incidence of holistic assessment (descriptive) in the three assessment systems.

***What factors affect the assessors' holistic assessment of the last years' apprentices' answer papers?***

How do the assessment systems affect the assessors' holistic assessment (explanatory)?

Of the various factors being investigated, the assessments systems will be central, and thus the vocational didactical principles on which these systems are based.

The theoretical framework has two main areas: vocational didactics and holistic assessment. The discussion about vocational didactics is based on both the international (Nordic and German) and the national research. This will be summarized in six perspectives for vocational didactics using the six assumptions developed by Gessler and Herrera (2015). It has not been possible to limit the theoretical framework for holistic assessment to one specific theoretical view, but emphasis has been on competence-based assessment (Wolf, 1995). The literature used has been systematized and processed with support in the didactic relationship model.

The study's research design and methods are developed on qualitative interviews conducted with thirteen assessors distributed within the three assessment systems and supported by the thinking of both "stimulated-recall" (Calderhead, 1981; Lyle, 2003) and "think-aloud" (Patton, 2015; Kuusela and Paul, 2000). Document analysis is a minor part of the data collection. The theoretical views of Charmaz's (2014), Strauss & Corbin' (1998) have supported the coding and categorisation of the empirical data. The results and discussions are presented in the respective categories. A further discussion have been summarized in four main dimensions.

### **The presentation of the study's central findings related to the three research questions:**

In the study's findings on the understanding of holistic assessment in a vocational didactical perspective, it became evident that little has been written on this subject. Through my development of a theoretical framework on holistic assessment, different perspectives emerged, and defined further by the vocational didactical framework. This is about highlighting a variant of VET assessment, where the assessor sees different aspects of a complex professional practice in a particular context by applying his experience, competence and judgment and assessing theses against the requirements for the profession.

Holistic assessment incorporates a view of assessment as one that promotes seeing an occupation as a professional whole. By highlighting holistic assessment in selected vocational didactical perspectives, opportunities, challenges and limitations are revealed. In particular, a vocational didactical affiliation and influence from working life will be central. The constant development in the trades is evident and affects singularly both the field of assessment and the field of vocational didactics, but also how these may be seen in context.

This again affects the understanding of what holistic assessment comprises. The confidence in the assessors who perform the assessments will be of significant importance here.

Findings show that the degree of holistic assessment amongst the assessors is high. This can be explained by the assessors' approach to the electricians' curriculum content, the incorporation of their own professional background, and how they applied assessment criteria in the assessment context. What promotes the holistic assessment is the ability to use the professional work-life experience in the assessment. The findings show that being part of an assessment network is central. In such a network, assessors can be informed, developed and updated. These aspects are important in electrical work, which is under constant development.

Findings also indicate that the form of the assessment system as well as current development in the electrical trade primarily affect the assessors' holistic assessment. Assessment systems affect the assessors by the design of the tasks, the structure of the assessment process and the given frameworks. The differences between the assessment systems are considerable and affects the assessors to different degrees. Findings show that the tighter and less flexible the system, the more it reduces the ability of the assessors to conduct a holistic assessment. The assessor's involvement and ownership of the assessment process is reduced and results largely in standardization and automation, where significant aspects of assessing holistic vocational competence will be overlooked. The assessment systems must adapt to the rapid development of the electricians' profession and give the assessors room to assess relevant professional competence. Central findings show that use of case-based tasks such as a work order, and the assessors' being given the opportunity to participate in assessment networks, contribute to the development and use of assessment criteria, have a positive impact on being able to make a holistic assessment.

In conclusion, the study indicates that there is a need for further research in several areas. These include assessment in other VET-programmes: different forms of assessment systems: development and application of goals and criteria: the importance of writing: and not least the development of vocational content.

# Innholdsfortegnelse

Forord .....	II
Sammendrag .....	III
Summary .....	VI
Innholdsfortegnelse .....	IX
Figurliste .....	XII
1. Innledning .....	1
1.1 Generell bakgrunn .....	1
1.2 Bakgrunn for HELVYRD-prosjektet og problemformulering .....	3
1.3 Studiens formål og forskningsspørsmål .....	5
1.4 Fag- og yrkesopplæringen i Norge .....	7
1.4.1 Struktur og bakgrunn .....	7
1.4.2 Elektrikerfaget .....	9
1.5 Presiseringer .....	11
1.6 Disposisjon av avhandlingen .....	12
2. Yrkesdidaktisk perspektiv .....	14
2.1 Innledende om yrkesdidaktikk og tilgrensende begreper .....	14
2.2 Yrkesdidaktiske betraktninger i en norsk og internasjonal kontekst .....	18
2.3 Diskusjon om yrkesdidaktiske perspektiver .....	24
2.4 Oppsummering yrkesdidaktiske perspektiver .....	30
3. Vurdering .....	31
3.1 Helhetlig vurdering .....	31
3.1.1 Sentrale begreper knyttet til vurdering .....	31
3.1.2 Didaktisk modell som støtte i strukturering av helhetlig vurdering .....	38
3.1.3 Oppfatninger om vurdering .....	40
3.1.4 Innhold .....	44
3.1.5 Rammer og kontekstuelle faktorer .....	45
3.1.6 Mål og kriterier .....	50
3.1.7 Metoder .....	55
3.1.8 Forutsetninger .....	59
3.1.9 Oppsummering teoretisk rammeverk for helhetlig vurdering .....	63
3.2 Forskning på vurdering i en internasjonal og norsk kontekst .....	64
3.2.1 Forskning på vurdering i en internasjonal kontekst .....	64
3.2.2 Forskning på vurdering i en norsk kontekst .....	71
3.2.3 Oppsummering om tidligere forskning .....	76

3.3 Oppsummering av vurderingsteoretisk rammeverk .....	79
4 Forskningsdesign og metode.....	81
4.1 Vurderingssystemene anvendt i HELVYRD-studien.....	81
4.1.1 KOMET-modellen anvendt i MECVET.....	81
4.1.2 Sentralgitt skriftlig eksamen.....	85
4.1.3 Fagprøven.....	89
4.2 Forholdet MECVET og HELVYRD.....	92
4.2.1 MECVET-designet og metodene.....	92
4.2.2 HELVYRD-designet.....	94
4.2.3 Hva svarer HELVYRD på kontra MECVET? .....	95
4.3 Generelle metodiske prinsipper for intervjuene i HELVYRD-prosjektet .....	96
4.3.1 Utvalg av informanter .....	98
4.3.2 Utvikling av intervjuguider og intervju spørsmål .....	101
4.4 Gjennomføring av intervjuer .....	104
4.4.1 Intervju av prøvenemndsmedlemmer til fagprøven .....	105
4.4.2 Intervju av ratere på MECVET besvarelser.....	106
4.4.3 Intervju av sensorer på sentralgitt skriftlig eksamen .....	107
4.4.4 Oppsummering og drøfting rundt gjennomføring av intervjuene .....	108
4.5 Dokumentanalyse.....	111
4.6 Etiske aspekter .....	112
4.7 Koding av empiri og utvikling av kategorier .....	115
4.8 Oppsummerende betraktninger om kvaliteten og validitet på design og metode .....	123
5. Presentasjon og diskusjon av resultater.....	128
5.1 Innhold.....	128
5.1.1 Vurderernes fokus .....	128
5.1.2 Diskusjon av vurderernes helhetlige vurdering av faglig innhold .....	136
5.1.3 Funn knyttet til helhetlig vurdering ved faglig innhold .....	138
5.2 Rammer og kontekstuelle faktorer .....	139
5.2.1 Vurderingssystemenes form og sammenheng.....	139
5.2.2 Utforming av oppgaver og tekst.....	142
5.2.3 Diskusjon av rammer og kontekster for helhetlig vurdering .....	144
5.2.4 Funn knyttet til helhetlig vurdering ved ulike rammer og kontekstuelle faktorer.....	149
5.3 Mål og kriterier .....	150
5.3.1 Kompetanse og faglige redegjørelser.....	151
5.3.2 Vurderingskriterier .....	153
5.3.3 Karakterenes betydning .....	155

5.3.4	Diskusjon om bruk av mål og kriterier i en helhetlig vurdering .....	159
5.3.5	Funn knyttet til helhetlig vurdering ved mål og kriterier .....	164
5.4	Metoder.....	166
5.4.1	Vurderingsstrategier.....	166
5.4.2	Vurderernes samstemthet .....	169
5.4.3	Diskusjon av vurderernes metodebruk i helhetlige vurderinger .....	172
5.4.4	Funn knyttet til helhetlig vurdering ved metoder .....	175
5.5	Forutsetninger .....	176
5.5.1	Fagperson som vurderer .....	176
5.5.2	Diskusjon om vurderernes forutsetninger for helhetlig vurdering .....	180
5.5.3	Funn knyttet til helhetlig vurdering ved forutsetninger.....	182
6.	Sammenfattende diskusjon og studiens funn: Helhetlig vurdering i et yrkesdidaktisk perspektiv	184
6.1	Mangfold og kompleksitet ved faglig innhold: Elektrikernes yrkeskompetanse .....	185
6.2	Vurderingssystemenes utfordringer ved en kompleks yrkeskompetanse.....	187
6.3	Vurderingspraksiser.....	191
6.4	Vurderernes profesjonelle erfaringsbakgrunn.....	197
6.5	Studiens funn.....	199
6.5.1	Syntetisering av studiens funn .....	199
6.5.2	Kunnskapssyn og helhetlig vurdering av yrkeskompetanse.....	201
6.5.3	Svar på forskningsspørsmålene.....	203
7.	Avslutning og forslag til videre forskning .....	206
7.1	Studiens bidrag.....	206
7.2	Refleksjoner om det metodiske ved studien.....	207
7.3	Refleksjoner om det teoretiske rammeverket .....	207
7.4	Forslag til tema for nye forskningsstudier.....	208
	Litteraturliste .....	211
	Vedlegg.....	223
	Vedlegg 1 Intervjuguider for vurderere i de tre utvalgte vurderingssystemene .....	223
	Vedlegg 2 Norsk samfunnsvitenskapelig datatjeneste AS (NSD) .....	229

## Figurliste

Figur 1. Sammenheng MECVET-prosjektet, norske vurderingssystemer og HELVYRD-prosjektet ....	5
Figur 2. Skisse for fag- og yrkesopplæring (kilde: vilbli.no). Normalløp er to år skole og to år bedrift (grønn pil, satt inn av meg).	8
Figur 3. Beskrivelse av elektrikerfaget (Kilde Udir.no)	10
Figur 4. Forskjeller mellom yrkesdidaktikk og fagdidaktikk. Hentet fra Johansson et al., referert i CEDEFOPs rapport <i>Vocational pedagogies and benefits for learners: Practices and challenges in Europe</i> , 2015, s. 23 (min oversettelse).	16
Figur 5. Dimensjonene for kompetansemodellen KOMET. Rauner (2013, s. 42)	23
Figur 6. Den didaktiske relasjonsmodellen med de ulike elementene basert på Bjørndal og Lieberg (1978, s. 135) og bearbeidet ved Hiim og Hippe (2009, s. 35)	38
Figur 7. Relasjonsmodellen knyttet til vurderingens form. Helle (2007, s. 201)	39
Figur 8. Viser systematiseringen av teori i ulike aspekter om helhetlig vurdering.	40
Figur 9. Kompetanse nivå og dimensjoner/kriterier KOMET-modell (Rauner, 2013, s. 48)	53
Figur 10. En oversikt på teoretisk rammeverk og sammenhengen mellom disse	80
Figur 11. Situasjonsbeskrivelsen og oppgavetekst (MECVET etter modell fra KOMET (Rauner, Haasler, Heinemann & Grollmann, 2009, s. 184)	82
Figur 12. Vurderingsskjema med åtte dimensjoner og 40 items (MECVET etter modell fra KOMET (Rauner et al., 2009, s. 182-183))	83
Figur 13. Løsningsrom (venstre) og eksempel på kompetanseprofil (høyre) (MECVET etter modell fra KOMET (Rauner et al., 2009, s. 185))	85
Figur 14. Informasjon til gjennomføring av forberedelsesdelen (venstre), og beskrivelse av situasjonen (høyre). (Kilde udir.no)	87
Figur 15. Informasjon til gjennomføring eksamen (venstre), utfyllende informasjon til de enkelte ordrene (midten) og skisse over plantegning (høyre). (Kilde Udir.no)	87
Figur 16. Sammenhengen mellom de utvalgte vurderingssystemene	92
Figur 17. Forskningsdesignet MECVET (hentet fra MECVET-søknad til Norges Forskningsråd (NFR))	93
Figur 18. Illustrering av hovedundersøkelsen og forholdet mellom det sekundære datasettet og primære datasettet. Det er selvsagt en glidende overgang mellom disse typene siden jeg har deltatt i innsamlingen av de sekundære dataene. Slik sett kunne forholdet og designet mitt være et eksempel på mixed methods der kvalitative og kvantitative data og analyse metoder integreres (Creswell, 2014, s. 14)	95
Figur 19. Oversikt på informanter fra de ulike vurderingssystemene	99
Figur 20. Gangen i gjennomføringen av intervjuene	103
Figur 21. Sammenhengen mellom forskningsspørsmålene og metoder for innhenting av data.	103
Figur 22. Gjennomføring av intervjuene med prøvenemndsmedlemmene ved fagprøven.	105
Figur 23. Gjennomføringen av intervjuene med ratere ved MECVET-testen.	106
Figur 24. Gjennomføringen av intervjuene med sensorene ved skriftlig eksamen	108
Figur 25. De ulike fasene i gjennomføring av intervjuene i de tre vurderingssystemene	109
Figur 26. Et eksempel (utsnitt) fra ratere i MECVET på hvordan kodingen har foregått ved bruk av Atlas TI.	117
Figur 27. Et utsnitt av initial koder og fordeling i de ulike vurderingssystemene.	118
Figur 28. Utsnitt av sorterte koder knyttet til vurderingssystem	119
Figur 29. Et av de første utkastene til kategorisering av koder.	120
Figur 30. En skisse på sammenhengen med teoretiske perspektiver og empiriske kategorier og underkategorier.	122
Figur 31. Forholdet mellom tema (koblet til teoretiske aspekter om helhetlig vurdering) og empiriske kategorier samt områder for de tre vurderingssystemene.	123



Figur 32. Faglige innholdskomponenter omtalt i studien.....	129
Figur 33. Oversikt over karakterer fra utvalgte eksamensbesvarelser.....	155
Figur 34. Oversikt over karakterer fra utvalgte fagprøver.....	156
Figur 35. Kompetanseprofiler fra elektrikerfaget på alle tre årene med elever til venstre, 1. årslæringer i midten og 2. årslæringer til høyre. ....	156
Figur 36. Oversikt på utforming og anvendelse av kriterier i de tre vurderingssystemene .....	165
Figur 37. Skissen viser de fire overordnede dimensjonene i den sammenfattende diskusjon, med utgangspunkt i funn om vurderernes helhetlige vurderinger fra tema med tilhørende kategorier samt de yrkesdidaktiske perspektiver. Avrundes med en syntetisering av studiens funn og en konkludering opp i mot forskningsspørsmålene.....	184



## 1. Innledning

Innledning til denne studien vil ta for seg en generell bakgrunn for tematikken og deretter en mer spesifikk bakgrunn for HELVYRD-prosjektet. Deretter presenteres studiens forskningsspørsmål. Så blir studien satt inn i en norsk kontekst med gjennomgang av fag- og yrkesopplæringsystemet. Deretter avrundes det med presiseringer og disposisjon.

### 1.1 Generell bakgrunn

I en OECD-rapport om norsk fag- og yrkesopplæring (Kuczera, Brunello, Field & Hoffman, 2008, s.13) anbefalte forfatterne både å etablere et nasjonalt rammeverk for vurdering i fag- og yrkesopplæringen og å styrke vurderingskompetansen til aktørene, spesielt i lærebedriftene og hos prøvenemnder. Tilsvarende anbefalinger ble gitt til flere andre land. I kjølvannet av dette igangsatte utdanningsmyndighetene flere prosjekter med mål om å øke kompetansen til lærerne og andre aktører innen fag- og yrkesopplæring når det gjelder å prøve ut både teoretiske og praktiske vurderingsformer i skole og bedrift. Flere utdanningspolitiske prioriteringer ligger bak satsningen. For det første har man fått et økt fokus fra myndighetens side på vurdering for og av læring (underveis og avsluttende) i en rekke norske offentlige utredninger (NOU) og stortingsmeldinger (Kunnskapsdepartementet, 2006, s. 31; 2011, s. 61; 2013, s. 33; 2016, s. 56; NOU 2008:18, 2008, s. 27; NOU 2014:7, 2014, s. 102; NOU 2015:8, 2015, s. 80; Utdannings- og forskningsdepartementet, 2004, s. 37). Disse utredningene omfatter både den formative og den summative delen av vurderingen. For det andre er vurderingsarbeid og vurderingsformer nå beskrevet i læreplanene etter Fagfornyelsen og dermed mer formalisert enn tidligere ettersom læreplanene er forskrifter. Prosessen med nye læreplaner har pågått samtidig som kompetansebaserte læreplanene i Kunnskapsløftet har gitt rom for lokale tilpasninger av opplæringen for den enkelte elev eller lærling. Dette gir både muligheter og utfordringer for de som skal utøve vurderingene, og vil kreve god innsikt i både faglig innhold og vurderingsformer for at de aktuelle kandidater skal få en rettferdig og relevant vurdering av egen kompetanse.

Den norske utdanningspolitikken gjenspeiler trender internasjonalt der elevens læringsutbytte har fått sterkere oppmerksomhet også innenfor fag- og yrkesopplæring. Det europeiske kvalifikasjonsrammeverket og PISA-studiene har bidratt til en intensivering av forskning og utviklingsarbeid om nye metoder og instrumenter for vurdering (Baethge & Arends, 2009, s. 8; Baumert, 2009, s. 25). Slike initiativ kan også sees på bakgrunn av reformer av fag- og

yrkesopplæring i flere land som bygger på kompetansebaserte modeller til tross for store forskjeller i struktur og kultur (Baumeler, 2019, s. 12).

En diskusjon av vurdering i fag- og yrkesopplæring vil kreve en utdyping av hva som skal vurderes. Selv om det har blitt vanlig å bruke begrepet «kompetanse» vil det også være fokus på «performance» eller oppgaveutførelse i en yrkessammenheng (Mulder, Weigel & Collins, 2007, s. 70). Mye av uenigheten kan tilbakeføres til ulike syn på kunnskap. Bagnall og Hodge (2017, s. 128) skiller mellom epistemologiske retninger (erkjennelsesteorier) som har preget diskusjonen av kompetanse. Det handler om en disiplinforankret epistemologi, konstruktivistisk epistemologi, emansipatorisk (sosiokulturell) epistemologi og en instrumentell epistemologi. Jeg kommer tilbake til disse i kapittel 3.1.1. Med ulike syn på kunnskap og kompetanse kan det reises flere spørsmål om hva som er helhetlig kompetanse og helhetlig vurdering. Yrkeskompetanse kan knyttes til det å mestre alle sider ved et yrke (Haasler & Erpenbeck s. 2008, s. 766), eller dens relevans vil være avhengig av personlig tilpasninger til et arbeidsliv (G. S. Haaland, Nore & Vagle, 2009, s. 218).

De ulike synene på kunnskap og kompetanse kan ikke på en enkel måte anvendes på feltet vurdering. Det er en utfordring å finne vurderingsformer som er like relevant på tvers av alle fag og vurderings situasjoner (Hiim & Hippe, 2001, s. 235). Med et helhetlig blikk på yrkeskompetanse vil man stille spørsmålet om det vil være vesentlig å se på vurderinger som også er helhetlige i sin form. Vil det innebære en større integrering av formativ og summativ vurdering (Thronsdén et al 2009, s. 29)? Vil en helhetlig vurdering bety at man kobler de ulike epistemologiene ved å fange lærlingenes yrkesdannelse på en måte og deres ferdigheter på en annen måte? Man må da stille seg spørsmålet om hva det innebærer å gjøre en helhetlig vurdering av en helhetlig yrkeskompetanse (Hager 2017, s. 209; Lum 2013, s. 1202). Skal vurdering av en kompleks yrkeskompetanse virke representativ må den ha en tett relasjon til det yrkeslivet og den yrkeskompetansen som skal utvikles og vurderes. Spørsmålet om vurdering i fag- og yrkesopplæring blir møtt med stor interesse fra næringslivet (G. S. Haaland et al., 2009, s. 215). Det uttrykte behovet beskrevet i NOU 2014:7 (2014, s. 10) og NOU 2015:8 (2015, s. 21) er at elever får en samlet kompetanse, noe som vil være sentralt for å utføre helhetlige vurderinger.

Spørsmålet om helhetlig vurdering av yrkeskompetanse kan diskuteres fra flere perspektiver. I denne studien vil jeg studere helhetlig vurdering av en yrkeskompetanse i et yrkesdidaktisk perspektiv. Vurdering av de lærendes yrkeskompetanse er en del av yrkesdidaktikken. Yrkesdidaktikken har en sentral plass i norsk fag- og yrkesopplæring og får mer og mer plass

i den internasjonale diskursen også (Rauner, 2013, s. 28; Gessler & Herrera, 2015, s. 153). Yrkesdidaktikken har sitt utgangspunkt i hvordan man kan legge til rette for opplæring som legger vekt på å møte utfordringene i et framtidig yrkesliv. Yrkesdidaktikken handler om å ha et tett forhold til yrkeslivet på den ene siden og på den andre siden binde sammen en hverdag som elev i skolen og lærling i bedrift. Men betydningen og forståelsen av begrepet skaper diskusjoner og brytninger som gjør at en videre forskning er nødvendig (Herrera, 2015, s. 167). Feltet er forholdsvis nytt. Det gir mulighet for denne avhandlingen å forme studien av helhetlig vurdering ved å ramme det inn i et yrkesdidaktisk perspektiv med støtte fra empirisk forskning av relativt ny dato. Det gir muligheter for nærmere å studere vurderernes vurderingshandlinger og de ulike faktorene som påvirker den helhetlige vurderingen av en kompleks yrkeskompetanse, noe som skal bidra til å sikre de krav som stilles i arbeidslivet. Dermed er det relevant å spørre hva helhetlig vurdering av elever og lærlinger i en norsk fag- og yrkesopplæringskontekst innebærer. Ikke minst er det viktig å få mer kunnskap om form og innhold i vurderingen. Kompetanse/yrkeskompetanse, helhetlig vurdering og yrkesdidaktikk er sentrale begreper i avhandlingen og utgjør grunnlaget for utviklingen av HELVYRD-prosjektet. Disse vil bli ytterligere utdypet i kapittel 2 og 3.

I neste avsnitt vil jeg ytterligere utdype motivasjonen for HELVYRD-prosjektet samt redegjøre for problemformuleringer som skal føre frem til avhandlingens forskningsspørsmål.

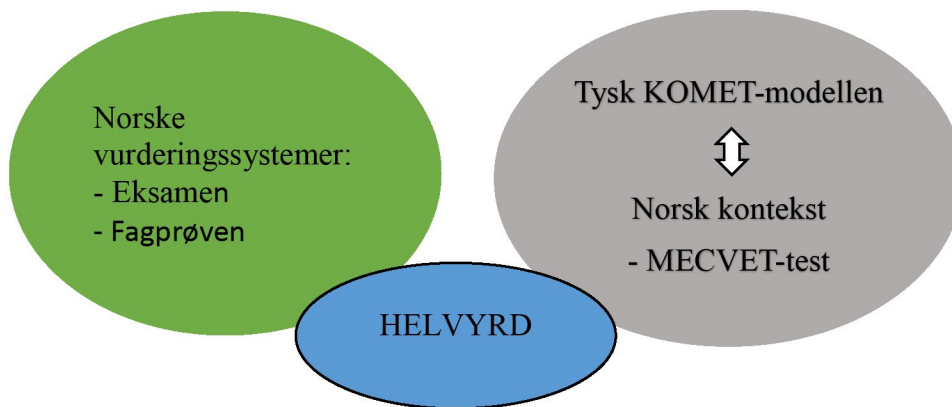
## 1.2 Bakgrunn for HELVYRD-prosjektet og problemformulering

I 2003 ble det i Tyskland lansert en ide om å lage en storskala undersøkelse for internasjonale sammenlignende studier etter modell av OECDs PISA-undersøkelse. Resultatene fra denne undersøkelsen ble presentert i 2006. Med basis i Universitetet i Göttingen ble det utviklet et konsept kalt VET-LSA, som ble utprøvd i flere land, deriblant Norge. Pilotprosjektet ble ikke videreført. Senere ble en annen modell for internasjonale storskala studier, Kompetenzdiagnostik im Berufsfeld Elektrotechnik (KOMET) med utspring ved IBB, Universitetet i Bremen, utviklet. Dette ble utgangspunkt for NFR (Norsk forskningsråd)-finansierte prosjektet Measuring Competence in Vocational Education and Training (MECVET). Ideen bak MECVET-prosjektet var å prøve ut KOMET-modellen i en norsk kontekst knyttet til utvalgte fag (tre fag: industrimekaniker, helsefagarbeider og elektrikerfaget) innen fag- og yrkesopplæringen. Flere elementer ved KOMET-prosjektets vurderingsinstrument måtte tilpasses samt videreutvikles. Det gjaldt spesielt vurderingskriterier til en norsk terminologi, utvikling av oppgaver med tilhørende

løsningsrom og prosedyrer for gjennomføring. Et av fokusområdene i utprøvingen var et behov for å gå dypere inn i hvordan vurdererne vurderte besvarelsene. Dette er noe HELVYRD-prosjektet griper fatt i for å forske på vurdererne som gjennomfører vurderingsarbeidet. Valget falt på å utforske dette i kun et av fagene: elektrikerfaget. KOMET-modellen (Rauner et al, 2013, s. 42) har som intensjon å kunne diagnostisere kompetanse for å kunne sammenligne på tvers av land. Det skilles mellom «performance assessment» og det å gjøre en diagnose på profesjonell kompetanse. Det vil i korte trekk si at praktiske ferdigheter ikke er en del av denne vurderingen, hvor det i stedet legges fokus på «cognitive performance dispositions» (s. 3). Sammenlignet med den tidligere nevnte PISA-undersøkelsen som tester de kognitive ferdighetene innenfor spesifikke fag, vil KOMET i større grad prøve å bygge oppunder vurderinger som er tett knyttet til et yrkesliv. Dermed etterstrebtes en mer helhetlig tankegang som skal ivareta ulike aspekter som blant annet problemløsning, organisering i arbeidslivet og faglig funksjonalitet. Spørsmålet er om denne tenkningen bygger oppunder en mer helhetlig tenkning enn andre storskalaundersøkelser som for eksempel PISA. Er det forskjell på å tenke helhetlig om vurdering tilknyttet yrkesfag kontra det enkelte fag? Tenker man også helhetlig når man vurderer en elevs mattebesvarelse? Hva ligger det av eventuelle forskjeller på å tenke helhetlig mellom yrker og et enkelt fag? I HELVYRD-prosjektet vil det først og fremst være sentralt å forske mer på hva helhetlig vurderingen innen yrkesfag vil innebære. Dermed vil vurdererernes vurderinger i MECVET-prosjektet være et utgangspunkt. Siden studien gjennomføres i Norge vil det være av interesse å trekke inn flere sentrale vurderingssystemer, for å kunne studere helhetlig vurdering i ulike kontekster innen fag- og yrkesopplæringen. Begrunnelsen for det er den kompleksitet og bredde som ligger i kompetansen til et yrke. Ulike vurderingssystemer med ulike tilnærminger kan gi interessante aspekter på hva helhetlig vurdering innebærer, og hvilke yrkesdidaktiske perspektiver som har påvirkning på vurdererens vurderinger. Dermed er begrunnelsen for å velge kun et fag, elektrikerfaget, i denne studien, et ønske om å studere vurderernes vurderinger mer inngående.

Dermed er HELVYRD-prosjektets utgangspunkt å studere vurderere i ulike vurderingssystemer og hvordan disse påvirker vurderernes helhetlige vurderinger i test og eksamenssituasjoner. Søkelyset rettes mot hvordan vurdererne, det vil si prøvenemndsmedlemmer, sensorer og ratere av kompetansetester, utfører sine vurderinger. For å studere vurdererens helhetlige vurderinger er det i HELVYRD-studien valgt ut tre vurderingssystemer innen vurdering av yrkeskompetanse. To av vurderingssystemene er integrert i den norske fag- og yrkesopplæringen hvor begge gjennomføres for lærlinger («high

stake»-tester som er viktige for kandidatens framtidige yrkesmuligheter). Den ene er sentralt gitt skriftlig eksamen og den andre er fagprøven. Det tredje vurderingssystemet er MECVET-prosjektet som har blitt prøvd ut i den norske fag- og yrkesopplæringen.



Figur 1. Sammenheng MECVET-prosjektet, norske vurderingssystemer og HELVYRD-prosjektet

Figur 1 viser sammenhengen i HELVYRD-prosjektet knyttet til både vurderingstesten i MECVET samt de to norske vurderingssystemene.

Spørsmål som kan stilles tilknyttet studien er om det finnes kriterier som styrer de valgene vurdererne gjør? Hvordan vil vurderernes egen faglige bakgrunn og kompetanse spille inn i vurderingsarbeidet? Legger vurdererne vekt på andre faktorer når de gjør sine bedømminger? Hvordan vil vurderernes tilknytning til deres eget fag og identitet spille en rolle i dette? Brukes de formelle styringsdokumentene og kriterier inn i dette arbeidet? Hvordan legger vurdererne opp sitt vurderingsarbeid? Og hvilken påvirkning har de enkelte vurderingssystemene på vurdererne? Spørsmålene bidrar i utviklingen av et grunnlag for studiens søkelys på vurderernes vurderinger og deres helhetlige vurderinger i et yrkesdidaktisk perspektiv.

HELVYRD-studien er en eksplorativ kvalitativ studie med noen få utvalgte informanter tilknyttet hvert vurderingssystem. Målet med studien er å utvikle kunnskap på et felt det er forsket lite på. Studien vil også kunne inspirere til utviklingsarbeid med nye vurderingsformer i fag- og yrkesopplæring.

### 1.3 Studiens formål og forskningsspørsmål

Med utgangspunkt i problemformuleringene beskrevet over er formålet med studien å belyse helhetlig vurdering i et yrkesdidaktisk perspektiv. Vekten vil ligge på omtale av rollen til

vurdererne i ulike vurderingssystemer og deres vurdering av yrkeskomptanse. Det vil også være sentralt å finne ut på hvilket grunnlag og med hvilke hjelpemiddel vurdererne gjør sine vurderinger, og hva som kommer fram i begrunnelsene for de valgene de gjør. Dette har ledet frem til formulering av tre forskningsspørsmål.

*-Hvordan kan vi forstå helhetlig vurdering i et yrkesdidaktisk perspektiv?*

Spørsmålet retter søkelyset mot et tema som i avhandlingen blir gjenstand for teoretisk drøfting og som gir en begrepsramme (*konseptuelt*) for å analysere det empiriske materialet. Spørsmålet skal svare på forholdet mellom helhetlig vurdering og yrkesdidaktikk. Målet er å belyse et yrkesdidaktisk perspektiv som kan påvirke vurderingen av en kompleks yrkeskompetanse. Helhetlig vurdering er sentralt og det vil være behov for å få frem en betydning av begrepet, slik at det kan diskuteres opp imot forståelsen av yrkesdidaktikken. Innen allmendidaktikk og fagdidaktikk omhandles også helhetlig vurdering og i kapittel 2.1 vil jeg kort diskutere de tilgrensende begrepene, noe som ytterligere vil tydeliggjøre mitt valg om å belyse det i et yrkesdidaktisk perspektiv.

*-I hvilken grad foretar vurdererne helhetlige vurderinger i de utvalgte vurderingssystemene?*

Dette spørsmålet skal besvares gjennom å beskrive innslaget av helhetlige vurderinger (*deskriptivt*) i de tre vurderingssystemene. Det å gjøre helhetlige vurderinger innebærer å forholde seg til en kompleks yrkeskompetanse hvor det vil være nærliggende å ha ulike tilnærminger. Vurderingssystemene vil bidra til at det vil være forskjellige måter å vurdere på. Dermed vil vurderernes helhetlige vurderinger studeres i ulike kontekster. Et viktig moment vil være å studere hvilke aspekter som legges til grunn for å kunne si noe om grad av helhetlig vurdering.

*-Hvilke faktorer påvirker vurderernes helhetlige vurdering av siste års lærlingbesvarelser?*

Hvordan påvirker vurderingssystemene vurderernes helhetlige vurdering (*forklarende*)? Av de ulike faktorene som skal undersøkes, vil vurderingssystemene som brukes innen fag- og yrkesopplæringen stå sentralt og de prinsippene disse bygger på. Det er sentralt å forstå de rammene et vurderingssystem utgjør og de prinsippene som ligger til grunn, slik at de handlinger en vurderer gjør kan diskuteres innenfor en bestemt vurderingskontekst. Samtidig vil det være av betydning å se de enkelte vurderingssystemene i lys av hverandre, for blant annet å studere om vurderingssystemene kan virke komplementære i vurderingen av en kompleks yrkeskompetanse. Det vil kunne bidra til å belyse betydningen av hva helhetlig vurdering innebærer.



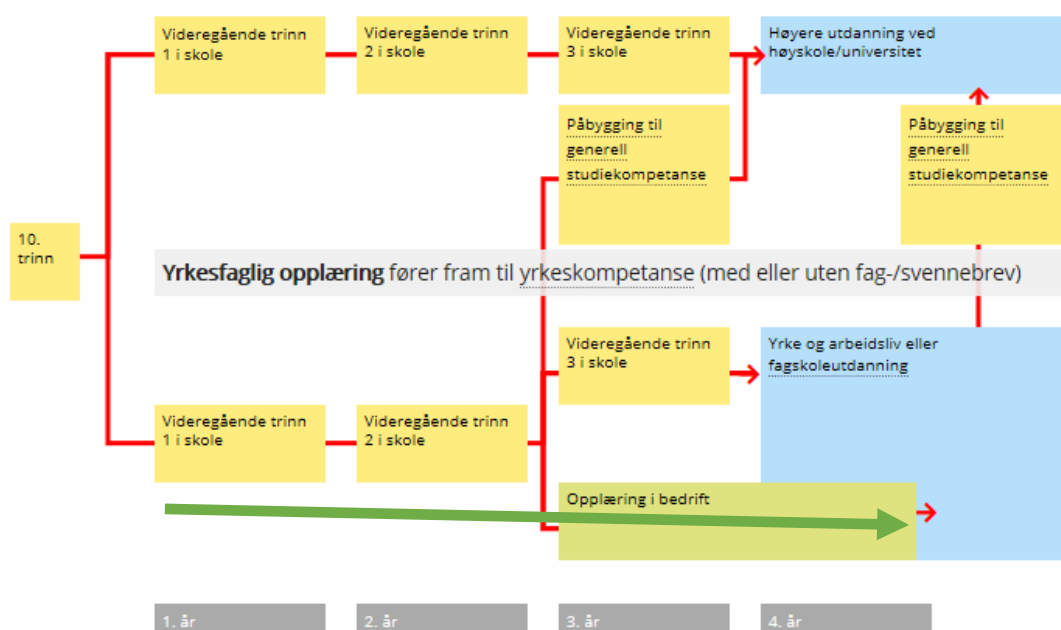
Gjennom forskningsspørsmålene søkes det å studere et felt som har en sentral betydning i fag- og yrkesopplæringen, og som skal bidra til å kvalifisere lærlinger til et arbeidsliv som har en stor samfunnsmessig betydning. Dermed skal forskningsspørsmålene kontekstualiseres gjennom en redegjørelse for den norske fag- og yrkesopplæringskonteksten.

## 1.4 Fag- og yrkesopplæringen i Norge

I dette avsnittet redegjøres det for den norske fag- og yrkesopplæringen gjennom å se på struktur og oppbygging samt på noen sentrale reformer som har hatt betydning for utviklingen de siste tiårene. Videre vil jeg presentere det utvalgte elektrikerfaget både slik det er presentert i formelle læreplaner og hva som påvirker faget i fremtiden. Dette vil være viktige faktorer tilknyttet vurdering.

### 1.4.1 Struktur og bakgrunn

Figur 2 viser en oversikt over hovedstrukturen i den videregående opplæringen i Norge med de mest vanligste alternative løpene. I den nedre halvdel av figuren vises det normale løpet for den yrkesfaglige delen. Fag- og yrkesopplæringen følger normalt en 2+2-modell med to år i skole (videregående trinn 1 (Vg1) og (videregående trinn 2 (Vg2)) og to år som lærling i (videregående trinn 3 (Vg3)) (grønn pil). I den norske modellen for fag- og yrkesopplæring er fag- og svenneprøvene på en måte yrkesopplæringsens kvalitetskontroll, og får oppmerksomheten fra alle parter som er involvert i fagopplæringen.



Figur 2. Skisse for fag- og yrkesopplæring (kilde: vilbli.no). Normalløp er to år skole og to år bedrift (grønn pil, satt inn av meg).

MECVET-testen benyttes på tre ulike steder, etter Vg2 i skolen, og ved både 1. års lærling og 2. årslærling på Vg3. Den skriftlige eksamen skjer halvveis i 2. årslærling og må bestås for å gå opp til fagprøven. Deretter er fagprøven helt på slutten av 2. årslærling. Den skriftlige eksamen er en 5 timers prøve. Oppgavene utvikles av en fagnemnd sentralt. Den sensureres lokalt av sensorer i de enkelte fylkeskommunene. De skal vurdere etter kompetansemålene i læreplanen. Karakterskalaen er fra 1 till 6 hvor seks er best. Fagprøven gjennomføres over flere dager og kan enten gjøres i bedrift eller på en prøvestasjon. Det er en prøvenemnd bestående av to prøvenemndsmedlemmer som vurderer. Den skal vurderes etter kompetansemålene i læreplanen. Det er en tredelt karakterskala (ikke bestått – bestått – bestått meget godt). Begge vurderingssystemene vil bli mer utførlig presentert i kapittel 4.1.2 og 4.1.3. MECVET-testen vil bli belyst gjennom den teoretiske redegjørelsen av KOMET-modellen i kapittel 2.2 og 3.1.6, og ytterligere beskrevet i kapittel 4.1.1.

Det er viktig å nevne at opplæringsmodellen presentert over har sitt utgangspunkt i to viktige reformer innen norsk fag- og yrkesopplæring, den første med Reform 94 i 1994 og den andre med Kunnskapsløftet i 2006. Reform 94 (Kirke-, utdannings- og forskningsdepartementet, 1996) innebar at all opplæring skulle føre fram til en avsluttende kompetanse, deriblant yrkeskompetanse. Den beskrivende kompetansen man skulle besitte etter endt opplæring bygde på et bredt kunnskapssyn. Dette innebar blant annet kunnskaper og ferdigheter, etiske verdier, kreativitet, selvstendighet i arbeidet, personlige kvaliteter og samarbeid og kommunikasjon. Inkludert i læreplanene var også områder som informasjonsteknologi, miljølære og internasjonalisering (under Fagspesifikke læreplaner punkt 4). I læreplanverket var den generelle læreplanen overordnet for alle fag. Kunnskapsløftet (LK06) var Stortingets behandling av St.meld. nr. 30 Kultur for læring (Utdannings- og forskningsdepartementet, 2004). Kunnskapsløftet var en videreføring av Reform 94 (Markussen, 2007, s. 97). Reformen innebar endringer av innhold, struktur og organisering. Noe av det viktigste var arbeid med grunnleggende ferdigheter (lese, skrive, muntlig, regne og bruke digitale verktøy), læreplaner med tydelige mål for kompetanse og ny fag- og timefordeling (Kunnskapsdepartementet, 2019).

Opplæringslova (Opplæringslova, 1998) og Forskriften til opplæringslova (Forskrift til opplæringslova, 2006) sammen med læreplanverket (Utdanningsdirektoratet, 2016a) utgjør de

sentrale dokumentene i norsk fag- og yrkesopplæring. Læreplanene er utarbeidet i tett samarbeid mellom partene i arbeidslivet og utdanningsmyndighetene (Utdanningsdirektoratet, 2017b). Hovedstrukturen i læreplanen inneholder et (1) formål med utdanningen, deretter en redegjørelse for (2) struktur på hovedområdene, så kommer (3) hovedområdene, som beskriver programfagene. Videre en redegjørelse for de (4) grunnleggende ferdigheter som omhandler det å kunne uttrykke seg muntlig og skriftlig, kunne lese, regne og å bruke digitale verktøy. Deretter kommer en beskrivelse av (5) kompetansemål. Helt til slutt kommer det bestemmelser knyttet til (6) vurdering (Utdanningsdirektoratet, 2013). I læreplanene i Kunnskapsløftet er det kompetansemålene som beskriver hva elever og lærlinger skal mestre etter endt opplæring (Utdanningsdirektoratet, 2009, s. 14). Kompetansemålene utgjør grunnlaget for opplæringen og vurderingen og innebærer at forhold som går ut over det rent faglige ikke skal trekkes inn i vurderingsarbeidet. I tilfellet elektrikerfaget er det i læreplan for Vg3 totalt 52 kompetansemål fordelt på tre programfag (Elenergisystemer, Tele-, data- og sikkerhetssystemer og Automatiserte systemer).

Meld. St. 28 om fag, fordypning og forståelse (Kunnskapsdepartementet, 2016) innebærer en fornyelse av Kunnskapsløftet, noe som vil medføre nye læreplaner for alle fag i 2020 samt en ny Overordnet del – verdier og prinsipper for grunnopplæringen (Kunnskapsdepartementet, 2017b). Fag- og yrkesopplæringen er i jevnlig utvikling for å tilpasse seg samfunnets behov, og Digitaliseringsstrategien for grunnopplæringen (Kunnskapsdepartementet, 2017a) er en del av det. I tillegg er det gitt signaler om at det overordnede målet er å sikre at elevene utvikler seg innen tre tverrfaglige tema; demokrati og medborgerskap, bærekraftig utvikling, folkehelse og livsmestring, samt at opplæringen skal preges av dybdelæring noe som er vesentlig for denne studiens tematikk.

#### 1.4.2 Elektrikerfaget

Mitt valg har vært å gå i dybden på ett fag, elektrikerfaget. Bemerke at dette faget har et avvik fra 2+2 modellen, hvor læretiden i bedrift utgjør 2,5 år (kommer mer tilbake til det i kapittel 4.1). Som en del av presiseringen vil det være nødvendig med en redegjørelse om innholdet og utøvelsen av yrket. I beskrivelsen av *elektrikerfaget* benyttes offisielle dokumenter som læreplaner (Utdanningsdirektoratet, 2013). Jeg har valgt å gjengi *Formålet* (s. 2) med faget i aktuell læreplan for elektrikerfaget (figur 3).

Elektrikerfaget inngår i et høyt teknologisk utviklet samfunn. Faget er knyttet til ulike arbeidsområder hvor elektrikerene har en viktig funksjon. I den forbindelse vises det til viktige egenskaper en elektriker må inneha. Det gjøres rede for et innhold i faget som lærlingen skal tilegne seg, men også hvordan lærlingen skal innorde seg et arbeidsmiljø og kunne kommunisere og samarbeide med andre. Videre beskrives arbeidsprosesser lærlingen skal jobbe etter. Dette skal utvikle en videre refleksjon som kan danne et grunnlag for livslange utviklings- og læringsprosesser.

Læreplanen er delt inn i tre hovedområder:  
Energisystemer, Tele-, data-, og sikkerhetssystemer og Automatiserte systemer. Disse har totalt 52 kompetansemål.

Et høyt utviklet teknologisk samfunn forutsetter sikre og velfungerende elektriske installasjonssystemer. Elektriske installasjonssystemer vil gjøre det mulig å oppfylle teknologiske krav til kvalitet, sikkerhet og komfort på viktige områder innenfor sjøfart, olje- og industriproduksjon, helsevesen, næringsvirksomhet og boliger. Kravene imøtekommes med godt kvalifiserte elektrikerer til å installere, kontrollere og vedlikeholde systemene og tilhørende utstyr. Gode tekniske og økonomiske løsninger vil tilfredsstille samfunnets teknologiske krav.

Opplæringen i elektrikerfaget skal utvikle lærlingens evne til å tenke helhetlig, og fremme system- og sikkerhetsforståelse og evne til omstilling. Nøyaktighet, kreativitet og løsningsorientering i utførelsen av arbeidet og økt bevissthet knyttet til dokumentasjon av opplæringen, skal også være sentrale faktorer. Elektrikerfaget skal øke bevisstheten hos lærlingen om lokale, nasjonale og globale miljøutfordringer knyttet til bedre ressursutnyttelse og en bærekraftig utvikling.

Opplæring i elektrikerfaget skal legge vekt på oppdaging, utprøving, innsikt og bevisste valg knyttet til energisystemer, tele-, data- og sikkerhetssystemer og automatiserte systemer og tilhørende utstyr. Forskrifter og kvalifikasjoner er sentrale elementer i elektrikerfaget og skal omfatte elsikkerhet, utførelse og reparasjon av elektriske anlegg. Bruk og forståelse av gjeldende regelverk om elsikkerhet samt det å kunne arbeide i henhold til fastlagte rutiner og prosedyrer skal være gjennomgående i opplæringen. Opplæringen skal fremme et godt arbeidsmiljø, selvstendighet og samarbeid med andre i og utenfor egen bedrift og fagområde, og den skal fremme evne til å kommunisere med brukere og kollegaer. Videre skal internkontroll, helse, miljø og sikkerhet, verdiskapning i samfunnet, internasjonalisering, serviceinnstilling og forståelse for bedriftens organisering ivaretas.

Opparbeidelse av evne til gjennomgående planlegging, dokumentasjon, system- og utstyrsforståelse og utvikling av læringsstrategier i virkelighetsnære arbeidsoppgaver, skal danne grunnlag for videre refleksjon, fordypning og spesialisering og fungere som byggesteiner for en livslang utviklings- og læringsprosess.

For kandidater som tar sikte på fagbrev som kvalifiserer for arbeid på norske skip, fiskefartøy og flyttbare innretninger, skal opplæringen gjennomføres i henhold til den internasjonale konvensjonen om normer for opplæring, sertifikater og vakthold for sjøfolk, 1978 (STCW) med endringer, herunder fagopplæring etter regel III/7 og sikkerhetsfamiliarisering etter regel VI/1.

Fullført og bestått opplæring fører fram til fagbrev. Yrkestittel er elektriker.

Figur 3. Beskrivelse av elektrikerfaget (Kilde Udir.no)

Innholdet i elektrikerfaget vil være i stadig utvikling, og digitalisering er en sentral del av fornyelsen av faget. Dette vil skje i et tett samarbeid med arbeidslivet. Meld. St. 27 om *Industrien – grønnere, smartere og mer nyskapende* (Nærings- og fiskeridepartementet, 2017, s. 65) redegjør for teknologidrevne trender der digitalisering og automatisering vil stå sentralt.

Alt dette vil ha betydning for innholdet i elektrikerfaget som er påvirket av nye tekniske innovasjoner. Myndighetene har faglige råd for fag- og yrkesopplæring. Når det gjelder elektro står det: «Vi skal identifisere trender og utviklingstrekk og sørge for at utdanningene våre er utformet slik at de dekker arbeidslivets, den enkeltes og samfunnets behov for kompetanse» (Utdanningsdirektoratet, 2017a, s. 2). I forbindelse med utviklingen av nye læreplaner har Faglig råd gjort en utviklingsredegjørelse med synspunkter på elektroområdet (Utdanningsdirektoratet, Faglig råd for elektro, 2016). Der er det redegjort for utviklingstrekk som vil ha betydning for elektrikerfaget. NELFO er en landsforening i Næringslivets Hovedorganisasjon (NHO) som blant annet omfatter elektrobedrifter. De skal bidra til å sikre et seriøst arbeidsliv og besitter kompetanse som skal sikre at yrkene får de rammene som trengs for å følge utviklingen. Det innbefatter også overgangen til et grønnere og mer bærekraftig samfunn. Elektrikerfagets utvikling vil bli ytterligere belyst i den sammenfattende diskusjonen i kapittel 6.

### 1.5 Presiseringer

De viktigste begrepene i oppgaven vil bli drøftet i kapitlene 2 og 3. Her vil jeg komme med noen presiseringer av mer teknisk karakter knyttet til den norske fag- og yrkesopplæringen.

Gjennomføringen av intervjuene måtte ta hensyn til gjennomføring av vurderinger, og det var vesentlig for studien å komme nærmest mulig selve vurderingsarbeidet. I HELVYRD-opplegget måtte det gjøres noen justeringer i fremgangsmåtene, noe som ga knapp tid i de innledende intervjuene med sensorene for eksamensbesvarelsene.

Når det gjelder vurderingssystemene, har det vært visse utfordringer med å tilpasse MECVET-testene til norske forhold. Dette gjelder spesielt oppgavene og kriteriene som ble brukt i vurderingsarbeidet. Erfaringene fra arbeidet viste at opplegget i den tyske KOMET-modellen ikke kunne bli direkte oversatt og brukt i Norge.

Når det gjelder informantene, vil de på tvers av vurderingssystemene bli kalt *vurdererne*. I hvert enkelt system vil de følge den offisielle betegnelsen:

- MECVET-test: *Ratere*

- Eksamen: *Sensorer*

- Fagprøven: *Nemndsmedlemmer*

De lærlingene som blir vurdert, blir i den offisielle sammenheng kalt kandidater, men jeg velger hovedsakelig å benytte *lærlinger*.

## 1.6 Disposisjon av avhandlingen

Kapittel 2 har som mål å utvikle et rammeverk for yrkesdidaktisk perspektiv knyttet til fag- og yrkesopplæringen. Yrkesdidaktikk med tilgrensende begreper behandles innledningsvis og deretter knyttet til diskurser i både norske og internasjonale miljøer. Dette vil sammenfattes i en diskusjon og utvikling av flere perspektiver om yrkesdidaktikken i kapittel 2.4. Disse vil bli trukket inn i den sammenfattende diskusjonen i kapittel 6.

Kapittel 3 er en gjennomgang av teori (kapittel 3.1) og forskning (kapittel 3.2) om vurdering som danner grunnlaget for utvikling av den teoretiske rammen om helhetlig vurdering.

Innledningsvis i kapittel 3.1 blir det en gjennomgang av sentrale begreper tilknyttet vurdering (kapittel 3.1.1). Systematiseringen av teorien og forskningen gjøres gjennom å anvende en didaktisk modell (kapittel 3.1.2).

I kapittel 4 redegjøres det for studiens forskningsdesign og metode. I kapittel 4.1 er det en gjennomgang av studiens utvalgte vurderingssystemer. I kapittel 4.2 beskrives sammenhengen mellom MECVET-prosjektet og HELVYRD-studien. I kapittel 4.3 beskrives utvalget av informanter og utviklingen av intervjuguider og spørsmål. Videre beskriver jeg i kapittel 4.4 hvordan intervjuene ble gjennomført i de ulike vurderingssystemene. Dokumentanalyse er en mindre del av datainnsamling og blir beskrevet i kapittel 4.5. I kapittel 4.6 blir ulike etiske sider ved HELVYRD-studien diskutert. I kapittel 4.7 tar jeg for meg fremgangsmåten i koding og kategoriseringsprosessen av studiens empiri og hvordan dette arbeidet har fremskredet. Avslutningsvis i kapittel 4.8 gjøres det noen betraktninger over forskningsdesignet sett opp mot noen valgte kriterier for kvalitative forskningsmetoder samt flere faser for validering.

I kapittel 5 er det en presentasjon og diskusjon av resultater i henhold til temaer og kategorier redegjort for i kapittel 4. Bakgrunn for å ha diskusjoner allerede i dette kapittelet er for å ha nærhet til empiri, teori og tidligere forskning. Funn blir oppsummert tilhørende vært tema. Dette utgjør utgangspunktet for den sammenfattende diskusjonen i kapittel 6.

Kapittel 6 er en sammenfattende diskusjon, organisert i fire overordnede dimensjon (kapittel 6.1 - 6.4). Kapittel 6.5 er rettet mot studiens overordnede funn. Den er delt i tre med en

syntetisering av studiens funn, kunnskapssyn og helhetlig vurdering og svar på forkningspørsmålene.

I kapittel 7 er en avslutning med forslag til videre forskning. Innledningsvis noen betrakninger om studiens bidrag. Deretter noen refleksjoner om både det metodiske og det teoretiske rammeverket. Avslutningsvis presenteres noen områder for videre forskning på feltet.

## 2. Yrkesdidaktisk perspektiv

I min studie vil jeg være opptatt av hvordan helhetlig vurdering kommer til uttrykk i vurderernes vurderinger i elektrikerfaget. I den sammenheng skal helhetlig vurdering belyses i et yrkesdidaktisk perspektiv. Det finnes ulike syn og oppfatninger blant forskere om betydningen av didaktikk begrepet og dermed «yrkesdidaktikk», og det blir sentralt å prøve å klargjøre noe av ulikhetene. Først blir det en innledende diskusjon om yrkesdidaktikk og tilgrensende begreper. Deretter blir det betraktninger om yrkesdidaktikk i en norsk og internasjonal diskurs. Basert på disse redegjørelsene vil jeg utforme beskrevne perspektiver om yrkesdidaktikk som skal bidra til å belyse helhetlig vurdering.

I arbeidet med å finne relevant litteratur om yrkesdidaktikk, utover egen tilgjengelig litteratur, har annen nordisk språklig litteratur blitt samlet gjennom søk i Oria, Academic Search Premier, ERIC, WebScience, SCOPUS og Google Scholar. Sentrale søkeord har vært «didaktikk», «fagdidaktikk», «yrkesdidaktikk» og «yrkespedagogikk». Også engelskspråklige ord som «vocational didactic» og «vocational pedagogy» har blitt anvendt.

### 2.1 Innledende om yrkesdidaktikk og tilgrensende begreper

Når det gjelder begrepet «yrkesdidaktikk», har det flere tilgrensede begreper som «didaktikk», «fagdidaktikk», «pedagogikk» og «yrkespedagogikk». Jeg vil kort gjennomgå disse begrepene innledningsvis med det målet å utdype betydningen av begrepet «yrkesdidaktikk».

Lyngsnes og Rismark (2014, s. 27) har i sin gjennomgang av didaktikkbegrepet vist til at fagområdet ikke er klart definert, og at det finnes både snevre og vide forståelser av begrepet. Herrera (2015, s. 162) hevder at ulike perspektiver og språk gjør at det blir et mangfold av definisjoner. Gudem (2011, s. 21) presiserer at begrepet didaktikk er mangfoldig, og kan ikke knyttes til bare ett felt. Ordet kan ha ulik betydning i den anglo-amerikanske litteraturen og i den kontinentale og nord-europeiske, selv om disse er i ferd med å tilnærme seg hverandre. Imsen (2009, s. 38) sier at didaktikkens innhold har utviklet seg med tiden. «Didaktikk handler om den opplæring, oppdragelse og sosialisering som skjer i skolen og i utdanningsinstitusjoner med et nærmere bestemt pedagogisk mandat». (Imsen, 2009, s. 37). Imsen mener at utviklingen av begrepet har gått fra et fokus på planlegging til å være mer utviklingsorientert, hvor skolens oppgaver og aktiviteter ses i en større samfunnsmessig sammenheng.



Gundem (2011, s. 75) uttrykker at «didaktikk» knyttes til pedagogiske sammenhenger som er komplekse, hvor forholdet mellom pedagogikk og didaktikk blir utfordret, og at forskjellen mellom begrepene «didaktikk» og «pedagogikk» utviskes. Gjennom de breie definisjoner som finnes av «didaktikk», som inkluderer opplæring, oppdragelse, sosialisering, nærmer en seg kjerneområder innen pedagogikk. Dette innebærer at didaktikken overtar områder som tidligere ble regnet som pedagogiske. (s. 78). Gundem (s. 78) viser til at det er en pågående kamp imellom tilhengere av de to begrepene og det vil være utfordrende å sette klare skiller imellom dem. Videre viser Gundem (s. 83) til samspillet mellom didaktikk og fagdidaktikk. Sistnevnte vil fremstå som mer konkret i forståelsen siden man anvender den direkte inn mot faget, hvor didaktikk, sett på som allmenndidaktikk, vil være mer abstrakt da den ikke er bundet til noen konkret undervisningsfelt. Men det vil være en vekselvirkning imellom disse to for å utvikle både allmenndidaktikk og fagdidaktikk.

Dette fører oss videre i diskusjonen om forholdet mellom begrepene «fagdidaktikk» og «yrkesdidaktikk». De oppfattes av enkelte som meningslike men av andre som to atskilte begreper (G. Haaland & Nilsen, 2013, s. 129; Hellne-Halvorsen, 2014, s. 23; Lyngsnes & Rismark, 2014, s. 35; Tarrou, 2004, s. 54). Tarrou trekkes fram som talsperson for det siste. Først og fremst brukes disse begrepene i lærerutdanningene generelt og yrkesfaglærerutdanningen spesielt. Videre blir det hevdet at bruksområdet for fagdidaktikk og yrkesdidaktikk er rettet inn imot skolen, den institusjonaliserte delen. Tarrou (2004, s. 56) sier at i planverket for både den yrkesfaglige delen i skolen og i lærerutdanningen «... brukes yrkesdidaktikkbegrepet normativt for å identifisere denne didaktikken som noe særegent, forskjellig fra didaktikken til de allmenne fagene, fagdidaktikken» (Tarrou, 2004, s. 56). Tarrou sier at yrkesdidaktikk «... forstått som det sentrale bindeleddet mellom yrkesfag i skolen, yrker i arbeidslivet og læringsprosesser og læringssystemer, utgjør kjernen i et yrkespedagogisk fagområde» (Tarrou, 2004, s. 56).

Studier rettet mot praktiske fag innen den yrkesfaglige delen av skolestrukturen kalles «yrkesdidaktiske». Det fagdidaktiske er rettet mer mot fag eller disipliner, og er mer knyttet til de enkelte fagene i skolens læreplaner (Tarrou, 2004, s. 56). Lyngsnes og Rismark (2014, s. 35) peker også på at fagdidaktikken er rettet inn imot skolefagene. Samtidig understreker de betydningen av å se tverrfaglig på de ulike fagene i skolen. Dette er et viktig element i yrkesdidaktikken, og har sammenheng med utviklingen i arbeidslivet. Johansson et al. viser i tabellform til forskjeller mellom yrkesdidaktikk og fagdidaktikk (subject-specific didactics). Figur 4 under er et eksempel på hvordan man kan illustrere noe av forskjellen.

Hovedforskjeller mellom yrkes- og fagspesifikk didaktikk	
Yrkesdidaktikk	Fagdidaktikk
Mest relatert til konstruktivistisk læringsteori	Mest relatert til kognitiv teori.
Undervisning og læring krysser flere grenser innen vitenskapens kunnskapsområder.	Ofte påvirket av kunnskaper fra universitetsfeltet, hvor det er et ønske å holde fagene separert.
Yrkesutdanning har sin didaktikk knyttet til arbeidslivet, ofte med et noe uskarpt kunnskapsområde.	Fagene er basert på rammer for den enkelte fagdisiplin.
Yrkesdidaktikk er handlingsorientert og hva som læres er ikke skilt fra nytteverdien.	Fokus på teori og kvaliteten på kunnskap.
Generelt så er det et problem som skal løses som krever både teoretiske og praktiske ferdigheter.	Basert på det didaktiske spørsmålet «hvordan?» noe som svares ved academia.
Det er ofte en taus og uformell læring inkludert i yrkesdidaktikk.	Kunnskapen kan ofte bli kommunisert verbalt eller i skriftlig form.
Yrkene er påvirket av forskjellige regler og reguleringer, noe som må være integrert i opplæringen.	Fagspesifikk didaktikk har ikke den samme forankringen.
Spesifikke krav fra arbeidsmarkedet og interesser har stor betydning, det samme gjelder arbeidskonteksten.	Kravene fra interesser gir rom for større frihet til tolkninger.

Figur 4. Forskjeller mellom yrkesdidaktikk og fagdidaktikk. Hentet fra Johansson et al., referert i CEDEFOPs rapport *Vocational pedagogies and benefits for learners: Practices and challenges in Europe*, 2015, s. 23 (min oversettelse).

Flere av sammenligningene av yrkesdidaktikk og fagdidaktikk i tabellen bekrefter at forskjellene handler om yrkesdidaktikkens tilhørighet til og påvirkning fra det arbeidsmarkedet man utdanner for. Samtidig viser (Nore, 2015, s. 192) til betydningen av digitalisering og ny teknologi, for eksempel sosiale medier. Yrkesdidaktikken må ta høyde for at flere aktører deltar i opplæringen. Nore viser til de hybride læringsarenaene som skaper flere elementer i læringsomgivelsene, noe som også kan utfordre grensene mellom kategoriene i tabellen.

Videre kan det settes et kritisk blikk på inndelingen mellom taus kunnskap på den ene siden og språklig formidlet kunnskap på den andre. Vitenskapsfilosofen Michael Polanyi mente at alle eksperter, også akademikere, trakk med seg «tacit knowledge» da «... they know more things than they can tell, knowing them only in practice, as instrumental particulars, and not explicitly, as objects» (Polanyi, 1962, s. 88). I tråd med dette vil jeg hevde at taus kunnskap gjør seg gjeldende innen områder hvor fagdidaktikken benyttes, også der skriftlig og muntlig

kommunikasjon er en vanlig formidlingsform. Digitaliseringen i arbeidsliv og skole vil utfordre delingen i tabellen siden nye verktøy krever en eksplisitt tilegnelse av begreper og prosedyrer. Dermed blir kommunikasjonen i yrkesfagene språkliggjort. Lyngsnes og Rismark (2014, s. 35) pekte på betydningen av å se tverrfaglig på skolefagene. Tabellen viser at dette kan utfordres, siden rammen må utvides utover den enkelte fagdisiplin. Skolefagene kan også nærme seg yrkesutdanningen. Studien til Lindberg (2003, s. 177) i den svenske yrkesskolen viste at i læring av et yrkesfag innebærer både praktisk lesing og kommunikasjon. I opplæringen brukes forklaring både som instruksjon og ikke minst når det gjelder å utvikle det yrkesmessige språket.

Når det gjelder skillet mellom «yrkespedagogikk» og «yrkesdidaktikk», sier Tarrou (2005, s. 197) at yrkespedagogikk er et vidt begrep. Det handler om en virksomhet som baserer seg på det erfaringsbaserte. Fotfestet er både arbeidslivspraksis og skolepraksis (det såkalte «doble praksisbegrepet»). Inglar (2009, s. 84) har i sin avhandling gjort rede for en definisjon og uttrykker at «... i en yrkespedagogisk læringsprosess vektlegges erfaringsbaserte og handlingsorienterte læringsformer, og læring gjennom samarbeid med andre» (Inglar, 2009, s. 84). Innholdet i læringsprosessen tar utgangspunkt i det som er sentrale yrkesfunksjoner. Ut i fra dette uttrykker Inglar følgende: «Yrkespedagogikk vektlegger kyndighetsutvikling der innholdet gis av sentrale yrkesfunksjoner, der pedagogisk teori tilpasses de aktuelle oppgaver, og arbeidsformene er induktive, handlings- og samarbeidsorienterte» (Inglar, 2009, s. 84). Med dette som bakgrunn har han definert «yrkespedagogikk» slik: «... de lærende skal systematisere de kunnskaper, ferdigheter og holdninger de allerede har med hensyn til læring, veiledning og undervisning, som grunnlag for helhetlig og praktisk-teoretisk kyndighetsutvikling» (Inglar, 2009, s. 324).

Beskrivelsene og definisjonen viser at det sentrale i helhetlig kyndighetsutvikling er knyttet til arbeidsoppgaver og yrkesfunksjoner. Behovet for en helhetlig utvikling av individet kunne i større grad vært understreket. Hellne-Halvorsen (2014, s. 20) har i sin gjennomgang av begrepet «yrkespedagogikk» knyttet definisjonen til rammeplanene for yrkesfaglærerutdanning og praktisk pedagogisk utdanning for yrkesfag (nettsider for planene). Hun har stilt spørsmålet om begrepet «yrkesdidaktikk» inkluderer det sentrale i «yrkespedagogikk», og om «yrkespedagogikk» muligens er erstattet av pedagogikk-begepet. Halvorsens spørsmål kan knyttes til Gudem (2011, s. 75) som påpekte at forholdet mellom pedagogikk og didaktikk ble utfordret, og at disse viskes ut.

I redegjørelsen om begrepet yrkesdidaktikk kommer det til syne at dette er et område som er i utvikling hvor skillelinjene til andre begreper stadig vekk flyttes. Men yrkesdidaktikk utgjør en kjerne i yrkespedagogikken og knyttes sterkt til yrkene det skal utdannes for, hvor sentrale arbeidsoppgaver, tverrfaglighet og helhetlig tenkning omkring opplæring er viktige perspektiver. Man kan si at det yrkesdidaktiske ligger nærmere det faglige innen yrket. I en slik kontekst vil det som særkjenner vurderingen i et yrkesdidaktisk perspektiv innebære ulike syn på hvordan vurderingen kan gjøres. Vurderingen må forholde seg til flere aspekter tilknyttet vurderingsprosessen, noe som igjen får betydning for hva som ligger i helhetlig vurdering. Ulike vurderingsteoretiske ståsteder vil kunne belyse ulike sider av det å vurdere en helhetlig yrkeskompetanse.

I neste avsnitt vil jeg se nærmere på hvordan disse spørsmålene er drøftet i den yrkesdidaktiske diskursen.

## 2.2 Yrkesdidaktiske betraktninger i en norsk og internasjonal kontekst

Jeg vil starte med å omtale den norske diskursen rundt yrkesdidaktikk. Deretter blir internasjonale synspunkter fra med våre nordiske naboer trukket inn. Siden den tyske KOMET-modellen er sentral i denne studien, vil de yrkesdidaktiske perspektivene som ligger til grunn for denne bli presentert. Deretter følger en presentasjon av to andre tyske forskere. Avslutningsvis gir jeg en oppsummering om videre drøfting av begrepet «didaktikk» og behov for mer forskning.

I den norske diskursen har Hiim og Hippe (2001, s. 31) i sin tolkning av «yrkesdidaktikk» utviklet en todelt definisjon av begrepet. I den første delen forstås begrepet som «praktisk-teoretisk planlegging, gjennomføring, vurdering og kritisk analyse av yrkesspesifikke utdanning-, undervisnings- og læringsprosesser». Videre er det lagt til: «... kritisk analyse og bruk av yrkesfunksjoner/yrkesoppgaver som grunnlag for læring» (Hiim & Hippe, 2001, s. 31). Fokuset ligger på funksjoner og oppgaver innen yrket som skal utgjøre grunnlaget for det som skal læres. Denne forståelsen deles av flere forskere (Brevik, 2014; G. Haaland & Nilsen, 2013; Inglar, 2009; Sannerud, 2005). Dette vil igjen få betydning for vurderingen av de lærende.

De nevnte forskerne har kommentert definisjonen og kommet med enkelte innvendinger. Inglar (2009, s. 65) reiser et spørsmål om de har tatt stilling til om yrkesdidaktisk kompetanse er kunnskap som er personlig eller om den oppstår og utøves når mennesket møter verden og

de handlinger som da oppstår. Hans tolkning er at den enkelte lærer besitter yrkesdidaktikk som en oppgave de jobber med, og at de ikke ser dette i en situert kontekst mellom elev og lærer. Sannerud (2005, s. 211) har pekt på at begrepet yrkesdidaktikk må defineres breiere og mener det er viktig å få med yrkeslæring som innbefatter arbeidsplassen. Brevik (2014, s. 32) trekker inn et aspekt om den forskjellighet som ligger i arbeidprosesser på skole kontra en bedrift. Tarrou (2005, s. 197) sier at arbeidsoppgavene i skole og bedrift ikke blir de samme. Sund (sitert i Nilsen og Haaland, 2013, s. 131) har forankret «yrkesdidaktikk» til de behov for kompetanse som gjelder i yrkesliv og samfunn, dessuten til læringsbehov hos den enkelte elev. Til forskjellen fra Sannerud (2005, s. 211) trekker de inn begrepet «kompetanse» og benytter elevperspektivet. Spørsmålet er om fokus da ligger mer på opplæring i skole enn i bedrift. Sannerud sin bruk av begrepet «lærende» kan virke mer dekkende for hele utdanningsløpet, inkludert opplæring i bedrift, enn det som kommer fram hos Sund.

Flere forskere trekker inn aspektet om yrkesdidaktikkens tilhørighet til arbeidslivet. Brevik (2014, s. 32) påpekte forskjeller mellom skole og bedrift, noe som er interessant med tanke på vurdering i skole og bedrift. Som en del av vurdering utvikles oppgaver for å teste kunnskaper og kompetanse. Lier (2017, s. 196) viser til utfordringer i det å gjøre arbeidsoppgaver om til læringsoppgaver, som i denne sammenheng kan ses på som vurderingsoppgaver.

Yrkesdidaktikken kan handle om å transformere en arbeidsoppgave til en læringsoppgave. En utfordring vil være å stille spørsmålet om hva en arbeidsoppgave er i utgangspunktet. Hvilken kompetanse er det man ser etter? De kompetansene som trengs og læres vil kunne være forskjellig fra ulike arbeidskontekster. Disse spørsmålene er sentralt å stille seg i utviklingen av oppgaver som skal anvendes i vurderingssammenheng. Om det vil være mulig å lage og å gjenskape en arbeidsoppgave som kan anvendes i en vurderingssituasjon vil kreve god innsikt i den aktuelle kontekst og den konkrete arbeidsoppgaven. Det fører oss over til diskusjonen om overføring mellom ulike arenaer.

I erkjennelsen av at transfer eller overføringsbegrepet blir noe snevert og ofte mekanisk forstått så bruker Lindberg (2003, s. 177) begrepet transformering. I et internasjonalt perspektiv konkluderer hun i sin studie at læring på skolen og i arbeidslivet ikke er så forskjellig som andre forfattere antyder. Begrunnelsen ligger i den hybride modellen mellom skole og arbeid hvor oppgaver som benyttes i skolen ligner de som gjelder i arbeidslivet. Forskjellen er at læring i skolen fokuserer på den sosiale læringen av yrket, mens læring i arbeidslivet fokuserer på produksjon. Samtidig i en nyere studie (Choy, Wärvik & Lindberg 2018, s. 345) redegjøres det for de kravene som bør ligge til grunn for at det skal skapes en integrert og god

opplæring imellom læringsarenaene både i skolen og arbeidslivet. Forfatterene diskuterer dette ut ifra fire områder: sosio-kulturelt perspektiv, enighet omkring læreplaner, interessenters roller og de lærendes forberedelser. Sistnevnte knyttes til Billett sin tenkning rundt best mulig forberedelse til praksis i bedrift (tabell 18.1, s. 362), og at det må foregå i forkant, underveis og i etterkant. Det er viktig at de lærende forstår det yrket de skal ut i praksis, kunne navigere og forstå læringsarenaen for å få best mulig læring og at det legges til rette for pedagogiske strategier for at læring skal skje tilpasset den lærende. I denne sammenheng vil Aarkrog (2010, s. 19) beskrivelse av transfer-begrepet være interessant med tanke på helhetlig vurdering.

Diskusjonen om transfer har mange kritiske røster, men Aarkrog viser til at transfer handler om å oppfatte identiske elementer eller likheter. Aarkrog uttaler at transfer må ses på som en transformasjon eller omdannelse, og at anvendelsen av det lærte må skje i en mest mulig lik situasjon (Aarkrog, 2010, s. 29). Oppmerksomheten mot undervisningens innhold gjør at transfer gjerne deles inn i *spesifikk* og *generell* (Aarkrog, 2010, s. 67). Det spesifikke handler om det konkret faglige, som læres og utøves. Det generelle handler om viten, eller ferdigheter som ikke er knyttet til en bestemt jobbfunksjon eller situasjon, men kan brukes i flere sammenhenger. Mens Brevik (2014, s. 32) trekker opp forskjellene mellom arbeidsprosesser i skole og bedrift, har Aarkrog (2005, s. 146) også sett nærmere på forholdet mellom det å lære i skolen og å lære i bedrift. Å utvikle en kvalifisert yrkesutøver innebærer å ha en vekselvirkning hvor man kan lære ulike ting på ulike steder og at dette utgjør en helhet. Hun presiserer at skolebasert utdanning også er viktig for å utvikle en kvalifisert yrkesutøver. Begge arenaer er også viktige for å utvikle spesifikk og generell kompetanse. Transfer er viktig imellom de to læringsstedene. Dette avhenger av hvilken praksissituasjon den lærende trenger for å ha et best mulig samspill med skolen. Betydningen av praksissituasjon og transfer er vesentlig når det gjelder å utvikle samspillet mellom praksis og skole i yrkesutdanningene. Dette kan variere mellom de ulike yrkesutdanningene (Aarkrog, 2004, s. 180). Knyttes dette til helhetlig vurdering, må de som skal gjennomføre vurderingene kunne kjenne igjen aktuell kompetanse for yrket. I et yrkesdidaktisk perspektiv vil transfer være viktig i utviklingen av systemer for vurdering hvor to faktorer gjelder. Det er utvikling av oppgaver og å gjenkjenne aktuell kompetanse under vurderingen.

Nilsson (1992, s. 109) bruker begrepet «fagdidaktikk» i tilknytning til valg av innhold og undervisningsmetoder. Dette ser han i et yrkespedagogisk perspektiv. Nilsson legger vekt på at den fagdidaktiske virksomheten skal basere seg på et spesifikt kunnskaps- og

innholdsområde. Han presiserer at en bør forankre dette området i den eksisterende tekniske utviklingen i arbeidslivet. Nilsson (2000, s. 246) uttrykker hvilken sentral plass arbeidsoppgavene har i yrkesdidaktikken:

Uppgifternas art är styrande för kunskapens karaktär. Det innebär att uppmärksamheten måste inriktas mot vad det är för art av uppgifter som förekommer inom yrkesområdet och hvilke krav uppgifterna ställer på den lärande människan. (Nilsson, 2000, s. 246)

Dette får konsekvenser for organisering av undervisning og innlæring. Undervisningen innen yrkesfag må ut fra dette ha et mer holistisk syn på både individ og på arbeidsoppgaver. I dette ligger det å se helhetlig på de arbeidsoppgaver som hører til det aktuelle yrket.

Johansson (2008, s. 139) peker på yrkesdidaktikkens sterke sammenheng med det arbeidslivet studentene skal forberedes til. Ut i fra et læreplansteoretisk utgangspunkt har han satt yrkesdidaktikk inn i en sammenheng med begrepene «formulerings»-, «transformerings»- og «realiseringsarena» (s.146) (parallell til Goodlad (1979, s. 348) sine fem læreplanteoretiske begreper ideologiske, formelle, oppfattede, gjennomførte og den erfarte). I begrepet «formuleringsarenaen» er det en tale om hvem som bidrar i diskusjonene om innhold. En av faktorene er at innholdet kan variere ut fra hvilken type arbeidsoppgave du skal gjennomføre i det aktuelle yrket. Dette betyr at det bør være en viss handlingsfrihet når det gjelder de nasjonale planene for opplæring, at disse kan tilpasses lokale behov. Dette fører diskusjonen over til transformeringsarenaen. Her skjer planleggingen av innholdet i en konkret undervisning. Yrkesfaglæreren må ta stilling til en rekke spørsmål knyttet til ulike rammefaktorer, blant annet hvor nær virkeligheten man kan komme med sin undervisning innen yrkesfag. Når alle rammefaktorene er tatt stilling til, går man over i realiseringsarenaen. Her blir undervisningen gjennomført, og elevene kommer i kontakt med sitt yrke. Her blir det en arena for hvordan man gjennomfører vurdering og som igjen har blitt påvirket gjennom formulering og transformeringsarenaen. Forskjellen mellom å lære på skolen og på arbeidsplassen er uklar. Opplæringen skal være nært tilknyttet det arbeidslivet man utdanner for. Dette vil prege vurderingen og den helhetlige tenkningen. Johansson stiller spørsmål om videreutvikling av yrkesdidaktikk, og om denne kan spesifiseres ytterligere relatert til det enkelte yrke. Er det nok å si «yrkesdidaktikk» eller må man inn på det spesifikke yrket for å konkretisere ytterligere den didaktikken som gjelder hvert yrke? (s. 155). Spørsmålet er viktig og vil bli diskutert videre i avhandlingen.

Som nevnt vil jeg belyse tysk yrkesdidaktikk siden KOMET-modellen er valgt. Der har man analysert arbeidsoppgaver og prosesser som grunnlag for å legge til rette undervisning. «Grounded work-based learning projects» viser til at læring og arbeid er nært relatert. Det overordnede med integrert formidling av kunnskap og praktiske ferdigheter er å utvikle helhetlig faglig ekspertise. Konseptet vil kreve at en implementerer «curriculum» slik at en retter seg mot arbeidsprosess-orientert undervisning og læring. Utfordringen har ligget i å utvikle vurderingssystemer som kan fange opp den reelle kompetansen som utvikles (Gessler & Howe, 2015, s. 233). KOMET-prosjektet eksemplifiserer denne tilnærmingen. Det har her vært et ønske om å utvikle en modell for testing av yrkeskompetanse, som gir en bedre innsikt i styrker og svakheter ved opplæringen. Rauner et al. (2013, s. vii) argumenterer med at modellen kan brukes i en sammenheng hvor utvikling av det didaktiske rundt undervisning er vesentlig. Oppmerksomheten retter seg mot overgangen fra en faginndelt disiplin i skolen til en utdanning bygd på utviklingslogikk kalt «læringsarenaer». Kjernen i KOMET-modellen er å utvikle virkelige case fra arbeidslivet som oppgaver for testing og læring (Rauner et al., 2013, s. 11). For å sikre den best mulige validiteten når disse skal nyttes i en test er det sentralt å ha relevante testoppgaver med sterk tilknytning til det utøvende yrket.

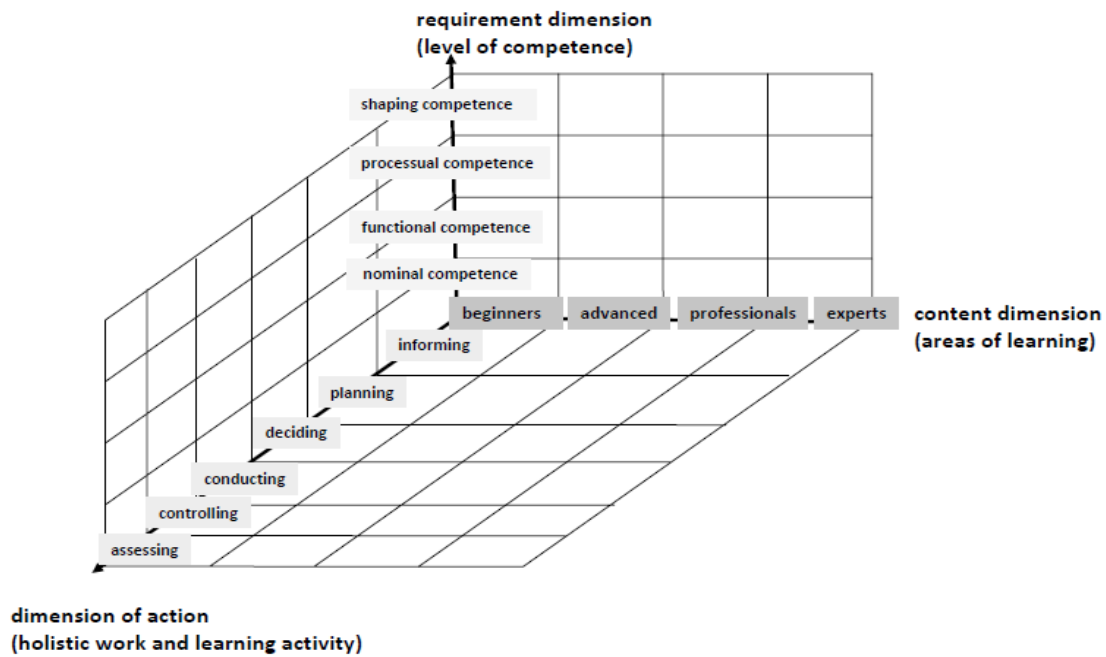
Når det gjelder å identifisere og analysere arbeidssituasjoner peker Rauner et al. (2013, s. 11) på tre faktorer: (1) Emnene, og det som gjenkjennes som et godt utført arbeid. (2) Verktøy, metoder og arbeidsorganisering. (3) Kravene til et profesjonelt arbeid. Logikken bak det å utvikle relevante oppgaver handler om å identifisere det karakteristiske ved profesjonelle arbeidsoppgaver, uttrykt som et didaktisk konsept, og som inkluderer fire læringsområder. Disse bygger på Dreyfus og Dreyfus' beskrivelse av utviklingen fra novise til ekspert:

- Knowledge for orientation and overview/orienting working tasks
- Integrated knowledge/systemic work tasks
- Knowledge of details and functions/problem-oriented, special work tasks
- Experiential and systematic in-depth knowledge/unpredictable work tasks

Rauner et al. (2013, s. 28)

Ut i fra dette er det utviklet en hovedmodell som grunnlag for tenkningen i KOMET. Den inneholder tre dimensjoner (figur 5) i forhold til arbeid og læring.





Figur 5. Dimensjonene for kompetansemodellen KOMET. Rauner (2013, s. 42)

Krav-dimensjonene (requirement) følger kriteriet om helhetlige løsninger, og er delt inn i fire nivåer. Disse går fra den nominelle nivået (laveste) til helhetlig «shaping» (høyeste) kompetanse som innebærer at den fulle komplekse forståelsen av de profesjonelle arbeidsoppgavene er realisert fullt ut. Innholds-dimensjonen (content) handler om fokus på hva som er karakteristisk for det som skal læres, og utgjør et grunnlag for å utvikle profesjonelle evner og talent for et yrke. I skalaen bygd på tenkningen til Dreyfus og Dreyfus () markeres innholdsdimensjonen fra novise, avansert nybegynner, kompetent utøver og profesjonell utøver. Aktivitets-dimensjonen (holistic work) handler om arbeidsprosesser med utvikling fra informasjon og videre planlegging, bestemmelse, gjennomføring, kontrollering og tilslutt evaluering. Sistnevnte om evaluering på en arbeidsprosessen kan være et utgangspunkt for å se på hvordan man kan gjennomføre vurdering i en opplæringskontekst. De yrkesdidaktiske dimensjonene i modellen vil kunne bidra til utviklingen av vurderingsformer.

De tre dimensjonene krav, innhold og arbeidsprosesser er sentrale i den yrkesdidaktiske tenkningen i KOMET-modellen. Målet er å beskrive et yrke ved sine karakteristiske arbeidsoppgaver gjennom analyser som grunnlag for å utvikle testoppgaver. Meningen er å måle profesjonell kompetanse, og ikke profesjonell handlingskompetanse (action competences). Dette betyr at den praktiske delen av yrket ikke vil være gjenstand for testing. Rauner skriver «What was evaluated was the conceptual competences in the sense of domain-specific cognitive performance dispositions» (Rauner et al., 2013, s. 3). Med testoppgavens

utgangspunkt i faktiske profesjonelle oppgaver vil det kunne være knyttet til verktøy, metoder og arbeidsorganisering og krav til et profesjonelt arbeid. Begrensningene vil være å kunne måle taus kunnskap, manuelle og kreative ferdigheter og sosial kompetanse (s. 13-16).

I redegjørelsen ovafor har jeg vist til nasjonale og internasjonale forfatteres syn på yrkesdidaktikk. De representerer et mangfold som viser at det trengs ytterligere drøfting av begrepet «yrkesdidaktikk», noe som vil kreve fortsatt samarbeid, deling og forskning. Gudem (2011, s. 19) viser til at didaktikkbegrepet har fått større utbredelse i engelsktalende land på grunn av et økende samarbeid internasjonalt. Ulike tradisjoner gir ulike føringer når det gjelder planlegging for undervisning og læring. Herrera (2015, s. 167) uttaler at “More is to be done in creating or continuing to develop a didactics that encompasses the complexities of VET considering its different levels and different settings. (Herrera, 2015, s. 167).

Forståelsen av didaktikk begrepet er viktig, men hva vi kaller didaktikken som utvikles for fag- og yrkesopplæring er ikke så viktig (Herrera, 2015, s. 167). Samtidig vil digitalisering og automatisering (Nærings- og fiskeridepartementet, 2017, s. 65) påvirke yrkers innhold og dermed yrkesdidaktikkens utvikling.

Ovenstående redegjørelser vil benyttes inn i neste kapittel for å utvikle perspektiver omkring yrkesdidaktikk.

### 2.3 Diskusjon om yrkesdidaktiske perspektiver

I diskusjonen om yrkesdidaktisk perspektiv har jeg valgt å bruke Gessler og Herrera (2015, s. 152) sine antagelser (assumptions) om og tilnærminger (approaches) til yrkesdidaktikken. I den videre redegjørelsen vil jeg anvende disse antagelsene og tilnærmingene som et grunnlag for å utvikle flere «perspektiver» (bruker dette begrepet heretter) om yrkesdidaktikk. De har gjort en analyse av det yrkesdidaktiske feltet i fem ulike land i Europa (Spania, Sverige, Danmark, Tyskland og Norge) noe som har ført fram til seks perspektiver. Analysen viser at det er behov for å diskutere løsninger ut fra de enkelte lands erfaringer med yrkesdidaktikk og utforming av yrkesopplæringen. Omtalen av perspektivene blir koblet til den ovenstående norske og internasjonale diskusjonen om yrkesdidaktikk. Perspektivene som er beskrevet må oppfattes som foreløpige og viser at det er behov for ytterligere forskning. Framstillingen vil likevel peke på områder som er vesentlige for å diskutere og utvikle yrkesdidaktikken. Selv om det eksplisitt ikke viser direkte til vurdering vil det allikevel være interessant å bruke omtalen av perspektivene som ramme for diskusjonen av helhetlig vurdering. Presentasjonens kronologiske rekkefølge er i denne omgang som i artikkelen.

*Det første perspektivet* handler om orientering mot læringsutbyttene og de krav som stilles i arbeidslivet. Dette innebærer at læringen skal utvikle kompetanse som tilfredsstillende gir krav. Samtidig må den lærende kunne være med på å utforme sitt eget arbeid og sin egen arbeidsplass (Gessler & Herrera, 2015, s. 153). I diskusjonen om yrkesdidaktikk kommer det fram hvor viktig det er legge vekt på hvilket arbeid den lærende skal utdannes for. Lyngsnes og Rismark (2014) stiller spørsmålet om hva en arbeidsoppgave kan være, og hvilken kompetanse man ønsker å utvikle. De viser til at dette vil avhenge av de varierende arbeidskontekstene. Lindberg (2003) og Aarkrog (2005) mener som nevnt at en vekselvirkning mellom skole og bedrift i en helhet utvikler den kvalifiserte yrkesutøver. Men gode forberedelser kreves for at den lærende skal få best mulig læring ute i praksis (Choy, Wärwik & Lindberg, 2018). Betydningen av å ha skole er viktig, men i en integrert sammenheng med det arbeidslivet man utdanner for. Tarrou (2005) viser til det «doble praksisbegrepet» som innebærer både skole og arbeidslivspraksis. Med opplæring bare i skole kan det stilles spørsmål ved hvor nær virkeligheten man kan komme (Johansson, 2008). Det vesentlige vil være å få til en helhetlig kyndighetsutvikling (Inglar, 2009) basert på læreprosesser tett tilknyttet sentrale yrkesfunksjoner. Sistnevnte støtter Rauner et al. (2013) og viser til hvor viktig det er å identifisere særtrekk ved profesjonelle arbeidsoppgaver når en skal utvikle relevante tester. Gessler og Howe (2015) uttrykker at «curriculum» må tilpasses arbeidsprosesser, og at opplæringen er deretter for å kunne utvikle en helhetlig faglig ekspertise. Utfordringen vil være å ha vurderingssystemer som kan fange opp den utviklede kompetansen. Samtidig er det et vesentlig spørsmål om hvem som skal bidra i diskusjonene om innhold (Johansson, 2008). Med en sammensatt yrkeskompetanse vil det ha stor betydning om hvilken røst som høres. Like viktig er bevisstheten om at innholdet kan variere ut fra arbeidsoppgavene som skal gjennomføres.

Ut fra det første perspektivet om fokus på læring vil utviklingen av den aktuelle kompetansen handle om tilknytning til det arbeidslivet man skal utdanne for, gjennom identifisering av yrkesoppgaver og bevissthet rundt faglig innhold. Både skole og arbeidsliv (og eventuelt andre arenaer) vil i en helhet kunne skape gode yrkesdidaktiske rammer. På den måten kan de tilfredsstillende gi kravet om en helhetlig kyndighetsutvikling relevant for det aktuelle yrket. Siden yrkeskompetansen er mangfoldig, må en samtidig ta hensyn til variasjon og tilpasning av faglig innhold i ulike kontekster. Dette vil få betydningen for hvordan man kan vurdere aktuell kompetanse og ivareta perspektivet om at man vurderer ut ifra hva som arbeidslivet krever.

*Det andre perspektivet* handler om at yrkesdidaktikken må reflektere den spesifikke kulturelle og historiske karakteren i et VET-system, iboende mål, verdier og normer, men også de forandringer som skjer i systemet. Ulike former for yrkesopplæring (statlig basert system, markedsorientert og en miks av begge) ligger til grunn i de utvalgte landene (Gessler & Herrera, 2015, s. 153). I denne sammenheng vil forståelsen av begreper være sentralt. Didaktikk-begrepet er i utvikling og feltet er mangfoldig (Gundem, 2011; Herrera, 2015; Imsen, 2009; Lyngsnes & Rismark, 2014). Samtidig er forståelsene i ferd med å nærme seg hverandre, noe som kan skyldes et større samarbeid på tvers av landene. Gundem (2011) viser til den breie forståelsen av didaktikk begrepet som samtidig utfordrer pedagogikk begrepet. Tarrou (2004) poeng om yrkesdidaktikkens særegenhet og den pågående diskusjon av begrepet gjør bevisstheten større. Samtidig utfordres den i takt med den utvikling som skjer i samfunnet for øvrig, men ikke minst i arbeidslivet hvor yrker både forsvinner og kommer til. Diskusjonen mot tilgrensende begreper bidrar til mer innsikt av begrepet «didaktikk». Grensegangen mellom fagdidaktikk og yrkesdidaktikk utfordres stadig (Tarrou 2004; Johanssons et al., figur 4). Det hevdes for eksempel at den tause kunnskapen (Polanyi, 1962) kan ha innvirkning på begge begrepene, spesielt hvor nye former oppstår med digitalisering (Nore, 2015). Når det gjelder yrkesdidaktikken kommer den sterke tilknytningen til arbeidslivet fram, noe som vil prege fag- og yrkesopplæringen. Samtidig tydeliggjør Hellne-Halvorsen (2014) yrkesdidaktikkens posisjon ved å hevde at begrepet yrkespedagogikk tilnærmer seg den mer allmenne pedagogikken. Hiim og Hippe (2001) sin definisjon av «yrkesdidaktikk» skaper en diskusjon blant norske forskere som kan sees på som en bevisstgjøring om den kulturelle utviklingen av begrepet. Et vesentlig spørsmål er om det kan snakkes om en generell yrkesdidaktikk, eller om den må sees i nær sammenheng med det enkelte yrke og den kontekst dette utøves (Johansson, 2008).

Det sentrale ved det andre perspektivet om den kulturelle og historiske karakteren ved yrkesdidaktikken er at den former seg over tid. Dette kommer til syne i diskusjonene om innholdet i begrepet yrkesdidaktikk og andre tilstøtende begreper, og utviklingen og forståelsen av definisjoner. Utbredelsen og tilnærmingen internasjonalt kan tildels tilskrives større samarbeid og interesse rundt fag- og yrkesopplæring. Yrkesdidaktikkens betydning og innhold er i stor grad prisdrevet den pågående kulturelle og samfunnsmessige utviklingen. Det kan bidra også i utviklingen av vurdering hvor ulike former og metoder anvendes hvor ulike syn på vurdering kan fremstå.

*Det tredje perspektivet* handler om horisontal struktur, at yrkesdidaktikken blir formet av de yrker som til enhver tid finnes. Denne forutsetningen er under konstant utvikling. Nye yrker kommer til, noen faller fra. Dessuten vil det være utvikling innen de enkelte yrkene (Gessler & Herrera, 2015, s. 156). Dette får konsekvenser for yrkesdidaktikken som bindeledd mellom skoler og yrker i arbeidslivet (Tarrou, 2004), noe som tvinger fram en stadig utvikling og tilpasning. Nilsson (1992; 2000) bekrefter dette ved å vise til behovet for å tilpasse undervisningen til den tekniske utviklingen, og fremmer et mer holistisk syn på den lærende og arbeidsoppgaven. Til Nilsson kan man stille spørsmålet om det holistiske går på personen eller den helhetlige arbeidsoppgaven. Det kan være utfordrende å skille disse da menneske og teknologi henger tett sammen i utøvelsen av yrket. Med Hiim og Hippe (2001) sin definisjon legges det vekt på en kritisk analyse og den sterke tilknytningen til arbeidslivet, hvor (Sannerud, 2005) peker på viktigheten av at yrkesdidaktikken også må innbefatte arbeidsplassen. I den sammenheng kan en sentral faktor i utviklingen av yrkesdidaktikken være å transformere reelle arbeidsoppgaver til læringsoppgaver med utgangspunkt i et aktuelt jobbdomene (Lier, 2017). Rauner et al. (2013) utvikling av realistiske oppgaver er basert på konkrete faktorer som kjennetegner et yrke. Disse må tilpasses fortløpende for å følge utviklingen og at man lærer av det. Her påpeker Aarkrog (2010) at det som læres et sted må brukes i et mest mulig likt sted en annen gang. Sett i en vurderingssammenheng vil det gjelde både utvikling av oppgaver men også å gjenkjenne kompetansen i yrket i selve vurderingen. I den sammenheng kan Johansson (2008) poeng om å se ulikheter mellom yrkene være interessant. Det kan ut fra dette være nødvendig å utvikle yrkesdidaktikken for det enkelte fag. En slik utvikling av yrkesdidaktikken krevet at nyanser fanges opp, noe som bare kan skje i et tett samarbeid. Aarkrog (2004) underbygger det ved å peke på at samspillet i transfer mellom skole og arbeidsliv kan variere mellom de ulike yrkesutdanningene.

Det tredje perspektivet om horisontal struktur handler om at yrkesdidaktikken formes av de til enhver tid eksisterende yrkene. Yrkesdidaktikken må være knyttet til et profesjonelt yrkesdomene for å ivareta relevansen av de transformeringer som skjer i relasjonen mellom opplæring og yrkesliv. Endringer kan også skje gjennom nye sammensetninger på tvers av eksisterende yrker. Vurderingen som skal foregå må følge etter og tilpasse seg de til enhver tid gjeldene yrkene slik at de tilpasses og får vurdert som er relevant.

*Det fjerde perspektivet* viser til en vertikal struktur, og at yrkesdidaktikken må forholde seg til de ulike faglige nivåene innen et yrke. Det kan være fra en ufaglært til en profesjonell utøver med høy grad av kompleksitet (Gessler & Herrera, 2015, s. 156). Rauner et al. (2013) viser til

utviklingen fra novise til ekspert (bygger på Dreyfus&Dreyfus) på innholdsdimensjonen og trekker også inn de krav som kreves sett fra nominell til shaping kompetanse. Han legger spesiell vekt på at det er ulike nivåer hos de lærende, og at dette må tilpasses i læring og testing. Aarkrog (2005) viser til den vekselvirkningen det er å ha opplæring i skole og bedrift, der praksissituasjonen sees som en viktig arena. I et transfer- perspektiv må det som skal læres bli brukt i en lignende situasjon, og på rett faglig nivå. Arbeidslivet er hierarkisk strukturert slik at en utdanning gir tilgang til et bestemt yrke, noe som innebærer at en fagarbeider må vite hva hun eller han ikke er autorisert til å gjøre. Samtidig kan viten eller mer generisk kompetanse (Aarkrog, 2010) ha betydning for hvordan man kan få utnyttet det i ulike sammenhenger og på ulike nivåer.

Det sentrale ved det fjerde perspektivet om vertikal struktur handler om skillet på de ulike faglige nivåene som finnes innen et yrke. Det er viktig å være bevisst på dette i opplæringssammenhengen, og at man i utøvelsen av yrket tilpasser arbeidsoppgavene den enkeltes kompetanse. Det vil ha direkte betydning for vurderingen hvor man må være bevisst det nivået som skal gjelde for aktuell utøvelse. Vurderingen må tilpasse seg det som er gjeldende for det aktuelle yrke.

*Det femte perspektivet* handler om at opplæringen kan være i ulike faser (temporal) av et langt arbeidsliv med både orientering, utdanning og videreutdanning (Gessler & Herrera, 2015, s. 156). Yrkesdidaktikken må tilpasses og utvikles med hensyn til utviklingsforløpets ulike faser. Johansson (2008) viser til at det er ikke noe enkelt svar på hva som er best av å lære i skole eller arbeidsliv, men Aarkrog (2005) framhever betydningen av å ha disse integrert. Samtidig peker Brevik (2014) på at det er forskjell på å lære i skolen og på en arbeidsplass, mens Lindberg (2003), som tidligere nevnt, påpeker at forskjellene ikke bør være store, men det krever struktur og forberedelser for å oppnå læring i praksis i bedrift (Choy, Wærwik & Lindberg, 2018). Her er et mangfold av meninger, noe som utfordrer yrkesdidaktikken når det gjelder å bruke flere arenaer for opplæring. Det er vesentlig å tilpasse seg de nødvendige behov når opplæring skal ivaretas gjennom et langt læringsliv. Om de samme yrkesdidaktiske prinsipper gjelder for unge som for voksne må diskuteres og prøves ut.

Ved yrkesdidaktikkens femte perspektiv om ulike faser (temporale), er det sentralt at opplæring må forholde seg til ulike arenaer gjennom et langt læringsliv, hvor yrkesdidaktikken må tilpasse seg de ulike forhold. Både skole og arbeidsplass har sine særskilte kontekster hvor tilrettelegging for læring vil innebære ulike utfordringer og

tilpasninger skal de imøtekomme de lærendes behov. Dette kan få betydning for vurderingen hvor man vil være bevisst på på hvor i et opplæringsløp en lærende er, noe som kan påvirke hva som vektlegges i vurderingen.

*Det sjette og siste perspektivet* handler om arbeidslivets naturlige globale endringer. Yrker og oppgaver kommer og går, noe som er viktig sett i et yrkesdidaktisk perspektiv (Gessler & Herrera, 2015, s. 157). Nilsson (1992) viser til den tekniske utviklingen i arbeidslivet. Den kan handle om det som skjer i den enkeltes daglige arbeid, men kan også handle om teknisk utvikling som endrer samfunnsstrukturer. Sund (i Haaland og Nilsen 2013) trekker fram betydningen av samfunnsutviklingen som vesentlig for den enkelte lærende. Ett eksempel er hvordan ny teknologi som robotisering av oppgaver kan virke inn på sysselsettingen og hvilke typer problemløsning som utføres av kyndige fagarbeidere. Dette vil være interessant i et internasjonalt perspektiv med økende globaliseringen av arbeidsmarkedet. Her kan ulikheter bli mindre, noe som har sammenheng med økt samarbeid på tvers av landegrensene. Dette fører til at yrkesdidaktikken har en nasjonal diskusjon om egne yrker, men også en internasjonal der den komparative dimensjonen trekkes fram når det gjelder yrker og opplæring. Dette gir et mangfold som vil føre til diskusjon om yrkesdidaktikkens betydning og innhold.

Yrkesdidaktikkens betydning i diskusjonen om fag- og yrkesopplæring innebærer å utdanne framtidens yrkesutøvere, noe som krever en stadig forskning og utvikling på feltet (Herrera, 2015). Det fører til at de involverte aktører utvikler kunnskap. Dette vil igjen skape arenaer for opplæring der vurdering av aktuell kompetanse skjer i en kontekst som oppleves som meningsfull for alle parter.

Det sjette perspektivet om arbeidslivets globale endringer vil ha direkte påvirkning på arbeidsliv og samfunnsstrukturer, noe som virker inn på yrkesdidaktikkens utvikling. Forskning og utvikling både nasjonalt og internasjonalt er viktig for å tilpasse seg og endre de behov som er vesentlige. Dette er nødvendig dersom yrkesdidaktikken skal holde tritt med de behov som stadig vekk avdekkes i sammenheng med fag- og yrkesopplæringen. I denne sammenheng må også vurderingen følge etter noe jeg innledningsvis viste til et stort fokus på å utvikle nye former for vurdering.

## 2.4 Oppsummering yrkesdidaktiske perspektiver

I gjennomgangen av begrepet «yrkesdidaktikk» var et viktig trekk at forståelsen og innholdet er i endring og den påvirkes av utviklingen i arbeidslivet. Den helhetlige kyndighetsutviklingen må sees i sammenheng med faktorer som transfer, digitalisering, taus kunnskap og mer eksplisitte kommunikasjonsformer, tverrfaglig samarbeid og fagets egenart og systemer som kan vurdere den komplekse kompetansen. I definisjonen av yrkesdidaktikk ligger Hiim og Hippe (2001) sin redegjørelse til grunn. Men et viktig moment i den sammenheng er Sannerud (2005) sin påpekning om det å ha et mer utvidet syn, hvor arbeidslivet også inkluderes. En vesentlig faktor ved yrkesdidaktikken er den sterke tilknytningen til det enkelte yrket, og i denne studiens tilfelle elektrikerfaget, noe som også er i tråd med Johansson (2008) sine betraktninger om behovet for å ha en yrkesdidaktisk tilpasning til det enkelte yrke.

Utviklingen av de seks yrkesdidaktiske perspektivene ble diskutert og utledet i kapittel 2.3 basert på redegjørelsene i kapittel 2.1 og 2.2. I prosessen ble Gessler og Herrera (2015) sine antagelser og tilnærminger til yrkesdidaktikken anvendt som en støtte for å utvikle perspektivene. Helhetlig vurdering vil belyses innenfor disse seks perspektivene: krav i arbeidslivet, historiske og kulturelle sider ved VET, forming av yrkene, faglige nivåer, ulike faser i opplæringen og tilslutt påvirkningen fra globale endringer. Perspektivene vil anvendes i den sammenfattende diskusjonen i kapittel 6.



## 3. Vurdering

Vurdering er en del av det yrkesdidaktiske feltet og i dette kapitlet skal det redegjøres helhetlig vurdering og forskning på feltet. Kapittel 3.1 omhandler det teoretiske rammeverket for helhetlig vurdering. Kapittel 3.2 er en gjennomgang av norsk og internasjonal forskning på feltet. Kapittel 3.3 inneholder en kort oppsummering av det valgte vurderingsteoretiske rammeverket (inkludert yrkesdidaktiske kapittel 2).

### 3.1 Helhetlig vurdering

I gjennomgang av litteratur om helhetlig vurdering og annen tilstøtende teori om vurdering har det vist seg at det ikke er mulig å avgrense seg til ett bestemt teoretisk ståsted. I utviklingen av et teoretisk rammeverk har jeg prøvd å systematisere litteraturen inn mot det sentrale i HELVYRD-studien. Som hjelp i systematiseringen har jeg tatt i bruk *Den didaktiske relasjonsmodellen*, noe jeg kommer tilbake til i kapittel 3.1.2.

I arbeidet med å finne relevant litteratur og tidligere forskning på feltet har databaser og søkekilder blitt anvendt, som Academic Search Premier, ERIC, WebScience, SCOPUS, Oria og Google Scholar. Sentrale søkeord om engelsk språklig litteratur om vurdering har vært «assessment», «judgment», «holistic assessment», «assessor», «rater», «vocational training/educational (VET)» og «grading». Utover egen tilgjengelig litteratur om vurdering har annen nordisk språklig litteratur blitt samlet gjennom søk i Oria. Sentrale begreper har vært vurdering, helhetlig vurdering.

Før jeg presenterer den utvalgte litteraturen vil jeg redegjøre for sentrale begreper i forskningen om vurdering.

#### 3.1.1 Sentrale begreper knyttet til vurdering

Sentralt tilknyttet vurdering er begrepet kompetanse og jeg vil først redegjøre for hvordan jeg vil forholde meg til det samt andre tilknyttede begreper. Videre redegjøres det for begrepet yrkeskompetanse, hvor hensikt er å frembringe forståelse av kompetanse tettere opp i mot fag- og yrkesopplæringen. Deretter runder jeg av med de utfordringer som ligger i innføringen av kompetansekonseptet og hva det innebærer for vurdering. Avsluningsvis vil andre tilknyttede begreper bli redegjort for.

Mye har vært skrevet om *kompetansebegrepet*, og det finnes flere historiske redegjørelser for bruk av begrepet og dets utvikling (Ellström, 1992; Eraut, 1998; Hager & Gonczi, 1996; Mulder et al., 2007). Det fremkommer ulike syn på hva som ligger i kompetanse, og definisjonen viser kompleksitet i innhold og forståelse av begrepet, noe som vil skape muligheter og utfordringer i vurderingssituasjoner. Begrepet kompetanse er en del av den offentlige diskusjonen i Norge. Eraut (1998, s. 137) påpeker at offisielle definisjoner kan gi en begrensning eller være for vag og at det kan være en utfordring. Den offisielle debatten holder seg til utviklingen av det nasjonale kvalifikasjonsrammeverket (NKR). I stortingsmeldinger og i offentlige utredninger har det vært en gjennomgang av kompetansebegrepet (Kunnskapsdepartementet, 2016, s. 28; NOU 2014:7, 2014, s. 9; NOU 2015:8, 2015, s. 18; NOU 2018:2, 2018, s. 14; NOU 2019:2, 2019, s. 17)

Ulike definisjoner har vært uttrykt blant annet med støtte i OECD og prosjektet DeSeCo (Definition and Selection of Competencies 2002). Kunnskapsdepartementet (2016, s. 28) bruker en definisjon som gjelder fagfornyelsen av Kunnskapsløftet: «Kompetanse er å tilegne seg og anvende kunnskaper og ferdigheter til å mestre utfordringer og løse oppgaver i kjente og ukjente sammenhenger og situasjoner. Kompetanse innebærer forståelse og evne til refleksjon og kritisk tenkning» (Kunnskapsdepartementet, 2016, s. 28).

Utdanningsdirektoratet viser til kompetanse forstått slik det er formulert i læreplanverket til Kunnskapsløftet:

Evnen til å løse oppgaver og mestre komplekse utfordringer. Elevene viser kompetanse i konkrete situasjoner ved å bruke kunnskaper og ferdigheter til å løse oppgaver. Det kan handle om å mestre utfordringer på konkrete områder innenfor utdanning, yrke og samfunnsliv eller på det personlige plan. (Utdanningsdirektoratet, 2016b).

Det er flere likhetstrekk mellom de to definisjonene, men det som skiller de to er Kunnskapsdepartementets vektlegging av refleksjon og kritisk tenkning. Det vil være naturlig i denne studien å forholde seg til definisjonen fra Kunnskapsdepartementet uttrykt i forbindelse med fagfornyelsen av Kunnskapsløftet, da den skal speile utviklingen innenfor fag- og yrkesopplæringen.

Nært tilknyttet *kompetanse* er begrepene *fagspesifikk* og *fagovergripende* kompetanse (NOU 2014:7, 2014, s. 55; NOU 2015:8, 2015, s. 19). *Fagspesifikk* handler om de kunnskapsområder skolefagene bygger på, og *fagovergripende* handler om det som er

relevant for flere ulike fag og kunnskapsområder. Kompetanse kan være generisk og overførbare til andre sammenhenger, men kan også være avgrenset til den aktuelle konteksten.

Forholdene mellom begrepet *kompetanse og kompetanser* (engelsk: competence and competencies) er diskutert. Eraut (1998, s. 133) påpeker distinksjonen mellom betydningen av de engelske begrepene ved vurdering og måling av kompetanse. Det stilles spørsmål om hvilke kompetanser som er inkludert i den bredere termen kompetanse, og hvordan disse delene utgjør en helhet. Kompetanser er mer oppdelt og oppstykket i mindre enheter, mens en samlet helhet utgjør en kompetanse. En oppdeling av kompetanse i flere kompetanser i en mer modulær form kan medføre en vurderingsform som i større grad kartlegger enkeltkunnskaper og ferdigheter (Wolf, 2001, s. 3). Samtidig uttrykker Hager og Gonczi (1996, punkt 2) at «... some degree of atomism in approaches to competence will be acceptable, provided that it is accompanied by a suitable degree of holism.» (Hager & Gonczi 1996, punkt 2).

Haasler og Erpenbeck (2008, s. 766) viser til *kvalifikasjonsbegrepet* og til distinksjonen mellom kompetanse og kvalifikasjon. Det sistnevnte tillegges den betydning at man besitter et sett kunnskaper og ferdigheter som kvalifiserer for de konkrete målene. *Kompetanse* blir dermed at individet handler under forhold som er usikre som inkluderer en mer helhetlig tenkning. Ellström (1992, s. 29) uttrykker at *kvalifikasjon* er krav knyttet til arbeidsoppgavens karakter, og etterspørres av en arbeidsgiver. *Kompetanse* utgår fra individet selv og det potensiale vedkommende har i arbeidet. Sannerud (2006, s. 228) har i sin gjennomgang uttrykt at begrepet kan oppfattes som smalere og mer snevert i forhold til kompetanse.

*Yrkeskompetanse* er utledet fra kompetansebegrepet. Haasler og Erpenbeck (2008, s. 766) sin forståelse av yrkeskompetanse er «... that competences are self-organized dispositions of complex adaptive systems i.e. individuals in particular» (Haasler & Erpenbeck, 2008, s. 766). Kompetanse er et potensial hos den enkelte til å handle adekvat i nye situasjoner. De besitter kompetanse som de selv greier å organisere slik at nye og andre arbeidsoppgaver kan løses. Begrepet innbefatter personlig kompetanse, teknisk/fag, metode kompetanse og sosial kompetanse, noe som utgjør en basiskompetanse. Forfatterne viser til at det blir utviklet lister med uendelige sub-kompetanser, men anbefaler at man holder seg til de tre (fire). Rauner et al. (2013, s. 6) viser til en modell for yrkeskompetanse knyttet til KOMET-modellen. Utenom de praktiske ferdighetene blir kunnskap om å legge til rette for handlinger (know that) vurdert, dessuten kunnskap om tolkning av handlinger (know how) og til slutt kunnskap om refleksjon

rundt handlinger (know why). Dette utgjør profesjonell arbeidsprosess-kunnskap. Sannerud (2006, s. 232) har i sin analyse av begrepet yrkeskompetanse tatt til orde for mer presist begrepsapparat som gir mulighet for en dypere analyse. I den sammenheng er det vesentlig å være bevisst på den forskjell som ligger i de ulike yrkene. For å kunne være presis på hva *yrkeskompetanse* er, trengs det at en undersøker yrke for yrke på gitte tidspunkt. Det argumenteres for at dette bør skje i nær tilknytning til den konteksten yrket blir utøvd i, noe som vil gi en mer fruktbar og presis beskrivelse. Rokkones og Fiskvik (2010, s. 191) uttrykker at yrkeskompetanse er nødvendig for å kunne utføre bestemte arbeidsoppgaver. Begrepet *arbeidsoppgaver* kan knyttes til det Sannerud omtaler som den konteksten yrket utøves i. G. S. Haaland et al. (2009, s. 215) sier at yrkeskompetanse er mer enn bare fagkompetanse, og at man i en yrkesutøvelse ofte må ha en tverrfaglig tilnærming til det problemet som skal løses. De har definert yrkeskompetanse som det å ha evne til å håndtere sammensatte utfordringer i et spesifikt yrke. I denne sammenheng viser Hiim og Hippe (2001, s. 54) til kompetanse som yrkeskunnskap, som er en mer bred forståelse av kunnskapsbegrepet, noe som innebærer at det teoretiske og praktiske knyttet til taus kunnskap og fortrolighetskunnskap må utgjøre en helhet. De uttrykker at «Dette innebærer at både intellektuelle, ferdighetsmessige, emosjonelle og holdningsmessige sider ved kunnskapen blir vektlagt» (Hiim & Hippe, 2001, s. 235). Mitt valg for denne studien er å forholde meg til begge begrepene: kompetanse og yrkeskompetanse. Førstnevnte i mer generelle redegjørelser og sistnevnte i direkte tilknytning til elektrikerfaget. I denne studien vil vurderernes forhold til kompetanse/yrkeskompetanse ytterligere bli operasjonalisert blant annet gjennom deres anvendelse av mål og kriterier fra både offisielle læreplaner og egenutviklede.

Innledningsvis i kapittel 1 ble det tatt opp at det er ulike synspunkter på kompetanse og dermed på hvordan kompetanse vurderes. Først trekker jeg frem betraktninger fra Mulder et al. (2007, s. 67) om innføringen av «The concept of competence» i utviklingen av VET i enkelte EU-land. I en gjennomgang av ulike tilnærminger til kompetanse finnes det flere definisjoner på bakgrunn av tvetydighet knyttet til læringsteorier og andre tilnærminger til læring. Dermed vil vurdering av kompetanse kunne være en krevende øvelse, spesielt hvis den blir knyttet til arbeidssituasjoner. Utfordringene med kompetanse og vurdering er summert opp i seks punkter (Mulder et al., 2007, s. 82). Det første punktet peker på at det oppstår en (1) ubalansert vektlegging ved vurdering av kompetanse. Det neste elementet viser at (2) vurdering av kompetanse kan oppleves mer frustrerende enn støttende for læring og utvikling. Deretter er det (3) en utfordring å avgjøre om en kompetanse er nådd eller ikke.

Utviklingen av kompetanser tar lang tid og noen av kompetansene vil kun benyttes etter utdanningen, og er dermed vanskelig å vurdere underveis i opplæringen. Videre påstås det at (4) formuleringen av kompetanse er for generell noe som igjen begrunnes med at effekten i vurderingen ikke vil være så tydelig. Vurderingen av (5) de lærendes utøvelse skjer ofte i en simulert profesjonell kontekst som ikke gir noen garantier for at personen vil være kompetent i en virkelig profesjonell kontekst. Det siste punktet handler om (6) eventuelle mangler på evne og troverdighet i analysene av kompetanseprofiler, og dermed muligheten til å vurdere om en person er kompetent nok. De seks punktene vil være interessante med tanke på de resultatene som vil komme i HELVYRD-studien. Spesielt når vurderingssystemene er bygd opp på ulike måter. Mulders betraktninger vil trekkes inn i den sammenfattende diskusjonen i kapittel 6. Tilknyttet de seks punktene, og kanskje spesielt det siste, viser Hager og Gonczi (1996, avsnitt før punkt 1) til at det å vurdere kompetanse innebærer å gjøre slutninger basert på bruddstykker av en kompetanse. Det er et vesentlig innspill i diskusjonen om kompetanse og vurdering, og noe som vil bli ytterligere berørt i den kommende teoripresentasjonen.

I tillegg til Mulder (2007) viste jeg innledningsvis kort til Bagnall og Hodge (2017, s. ) sine fire epistemologiske retninger (erkjennelsestrier) med ulikt syn på kunnskap. Det er et forsøk på en innramming som knyttes til utdanning og dermed også vurdering. Først kort presentere disse. Det første omhandler en (1) disiplinforankret epistemologi som var knyttet til skolefag og yrker. Innenfor denne tilnærmingen har kognitive teorier vært dominerende der individuelle eksamener er den vanligste vurderingsformen. Den (2) konstruktivistiske epistemologien legger vekt på personlig utvikling og mestring i autentiske omgivelser noe som innebærer en vurderingsform som legger vekt på personen som helhet. (3) Emansipatorisk, eller sosiokulturell epistemologi, ser yrkesutøverens utvikling i ulike kontekster. Den siste er (4) instrumentell epistemologi som ligger bak mye av tenkningen om vurdering knyttet til konkrete mål.

Slik de fire epistemologiske retningene Bagnall og Hodge (2017, s. 128) har utledet vil min antakelse være at alle fire kan være representert i denne studien om helhetlige vurderinger innen fag- og yrkesopplæring. Et *instrumentelt* syn kan innebære en vurderingsform som legger vekt på praktisk håndtering av enkeltoppgaver. Man kan oppfatte yrkeskompetansen som multi-dimensjonell «performance» men dette innebærer ikke nødvendigvis at vi kan omtale yrkeskompetansen som helhetlig (Lahn et al 2019, s. 147). I tråd med det *sosiokulturelle* synet er det ikke tilstrekkelig at yrkeskompetansen gjenspeiler de behovene som gjelder i arbeidslivet (Haaland et al., 2009, s. 215). Kompetansen må understøtte

tilpasning til yrker i endring hvor kritisk holdning og kreativ mestring av nye omgivelser er viktig. En yrkeskompetanse vil til en viss grad ha forankring i et *disiplinforankret* syn for eksempel på elektrofaget, men også inneholde andre kompetanser som går på tvers av fag – grunnleggende ferdigheter, sosial kompetanse og mer. I likhet med det *konstruktivistiske synet* betrakter Hiim og Hippe (2001, s. 243) yrkeskunnskap som en udelelig helhet og den erfarne utøveren utvikler et helhetlig blikk (se også Dreyfus & Dreyfus).

Eksemplifiseringen viser hvordan de fire epistemologiske retningene kan være representert gjennom ulike syn på vurdering av yrkeskompetanse. Siden min studie skal forske på tre varianter av vurdering, bygger det oppunder mitt valg om å studere flere sider av en vurderers helhetlige vurderinger (ref innledningsvis i kapittel 3.1). Derfor utvikles et teoretisk rammeverk gjennom å systematisere det i ulike aspekter ved helhetlig vurdering (se redegjørelse neste kapittel 3.1.2). Målet er å få en større forståelse av hva helhetlig vurdering kan innebære, blant annet ved å se i hvilken grad det gjøres og hva som påvirker vurderingene. Jeg kommer til å bruke Bagnall og Hodge (2017) sin inndeling av de fire epistemologiske retninger for både å plassere ulike teoretiske bidrag om helhetlig vurdering i kapittel 3 og mot studiens funn i kapittel 6.5.2. Sistnevnte kan mer ses på som epistemologiske hverdagsoppfatninger (Schommer-Aikins, Unruh og Morphew, 2017, s. 142), som kan identifiseres i de respektive vurderingssystemene og vurderernes synspunkter.

Avslutningsvis vil jeg redegjøre for tre begreper. Begrepet *vurderingssystemer* «... referer til den prosedyren eller virksomheten som er utformet for å samle inn informasjon om kunnskap, ferdigheter eller holdninger til en elev – forstått som den lærende – eller til en gruppe av elever». (Dale & Wærness, 2007, s. 105). Vurderingssystemer bør legge vekt på å motivere de lærende i opplæringen i stedet for å fokusere på de nederlag som kan oppstå (Dale & Wærness, 2007, s. 149). Dette sees i lys av det å danne et nasjonalt vurderingssystem, og tenkningen kan knyttes til min forståelse av hva som beskrives som vurderingssystem i HELVYRD-studien.

I henhold til Scriven er forholdet mellom begrepene *formativ* og *summativ* vurdering ofte satt opp som dikotomier (gjengitt i Thronsen, Hopfenbeck, Lie & Dale, 2009, s. 28). Formativ vurdering har som en klar hensikt å fremme læring og undervisning, og «This feedback helps to regulate teaching so that the pace of moving toward a learning goal is adjusted to ensure the active participation of the students» (Harlen, 2007, s. 19). Den skal utvikle den lærendes læring ved å gi relevant tilbakemelding på hvordan den lærende ligger an i forhold til målene.

Tanggaard og Elmholdt (2008, s. 98) beskriver formativ vurdering som det å være proaktiv, og benyttes for å forbedre og utvikle en aktivitet. Summativ vurdering har som mål å informere samfunnet om de resultater og den sluttkompetanse den lærende har oppnådd (Thronsen et al., 2009, s. 29). Dette er mer en reaktiv vurdering som legger vekt på resultatet av en handling snarere et handlingens læringspotensial (Tanggaard & Elmholdt, 2008, s. 98). Innholdet i vurderingen kan ta utgangspunkt i et arbeid, spesielt utviklede oppgaver, tester lagd av de som underviser, eller eksterne tester (Harlen, 2007, s. 21 og 23). I en rapport fra forskningsorganisasjon EARLI: *European Assosiation for resaearch on Learning and instruction* (gjengitt i Thronsen et al., 2009, s. 29) diskuterer man at summativ vurdering handler om at den har blitt lagt for stor vekt på. Det pekes på behovet for mer moderne vurderingssystemer hvor både formativ og summativ vurdering utvikles og integreres i utdanningen. Målet vil være å øke autensiteten og fleksibiliteten og minske prinsippet «teaching for the test». I de offisielle dokumenter er *summativ og formativ vurdering* (slutt- og underveisvurdering) noe som gjøres underveis eller til slutt, også nevnt som vurdering for og av læring. Skillet kan oppleves mindre klart enn man ofte finner gjengitt i litteraturen (Kunnskapsdepartementet, 2011, s. 61-64; 2013, s. 33; 2016, s. 57-59; Utdannings- og forskningsdepartementet, 2004, s. 37-39).

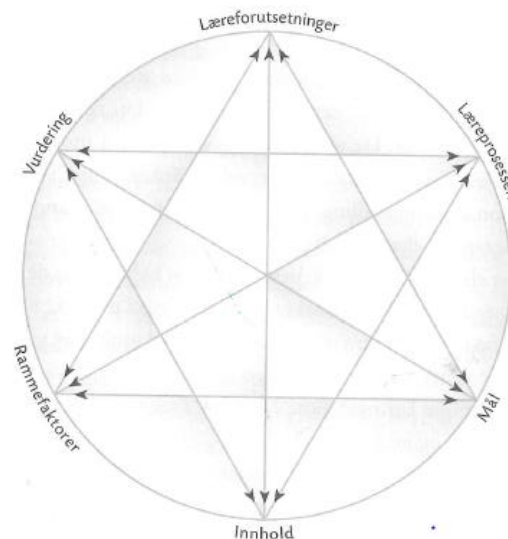
*Norm- og kriteriebasert* vurdering er to prinsipper omtalt av flere forfattere (Carlgren, Forsberg & Lindberg, 2009; Dale & Wærness, 2007; Ecclestone, 2010). Når det gjelder synet på vurderingsmetoder har det vært et skifte fra det norm-baserte til det mer kriterie-baserte. Norm-basert innebærer at sammenligning mellom elevene gjøres på grunnlag av fordelingen elevene imellom (Dale & Wærness, 2007, s. 146). Ecclestone (2010, s. 338) tillegger den kriterie-baserte vurderingen å kunne måle «the real-life skills» og at det i større grad møter de krav som stilles i samfunnet. I dette systemet er det lagt vekt på å fremme de forventede kompetanser og definerte kvalifikasjoner som kreves på ulike nivåer, hvor man med en mer valid og autentisk metode prøver å måle utøvelsen i forhold til de kriterier som er satt opp. Carlgren et al. (2009, s. 90) trekker også inn en tredje form, domene-relatert vurdering, og viser til ulike betydninger med paralleller til norm- og kriteriebasert. Et vesentlig aspekt ved de ulike formene er at det kan oppleves utfordrende å være klar på innhold knyttet til den enkelte, og det anses som viktig å være klar over de innbyrdes betydningene de ulike begrepene har (Carlgren et al., 2009, s. 90). I de offisielle dokumenter er *norm- og kriteriebasert* vurdering tatt opp i forbindelse med karaktersetting og at læreplanene har

kompetansemål de skal forholde seg til (Utdannings- og forskningsdepartementet, 2004, s. 39).

Etter gjennomgangen av ulike sentrale begreper vil det være sentralt å ha med seg betydningen av kompetanse og yrkeskompetanse og hvordan de epistemologiske retningene vil bidra i teori som presenteres og belyser helhetlig vurdering. Men først en redegjørelse om hvordan det gjøres ved hjelp av en didaktisk modell.

### 3.1.2 Didaktisk modell som støtte i strukturering av helhetlig vurdering

I min behandling av helhetlig vurdering har jeg blitt inspirert av didaktiske utredninger, funn fra empiri, søk i litteratur og forskning på feltet. Dette har ført til at jeg bruker *Den didaktiske relasjonsmodellen* (figur 6) med utgangspunkt i Bjørndal og Lieberg (1978, s. 130) og bearbeidet av Hiim og Hippe (2009, s. 37) når jeg utvikler en ramme for å systematisere aktuell teori om helhetlig vurdering og annen tilstøtende teori om vurdering.



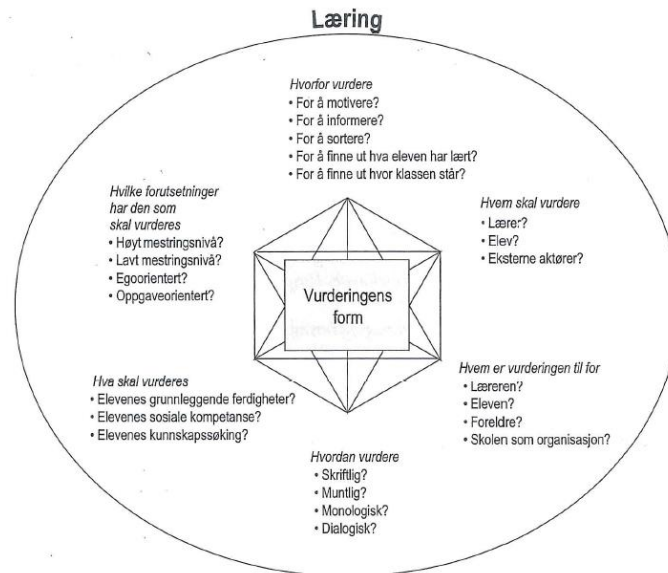
Figur 6. Den didaktiske relasjonsmodellen med de ulike elementene basert på Bjørndal og Lieberg (1978, s. 135) og bearbeidet ved Hiim og Hippe (2009, s. 35)

Først en kort kommentar til modellen. De didaktiske elementene må sees i en sammenheng, det vil si at endringer i ett element vil få innvirkning på de andre elementene.

Relasjonsmodellen har både deskriptive og normative funksjoner, og den anvendes både i planlegging og analyse av undervisning. Elementene i modellen skal oppfattes som sideordnet (Vestøl, 2008, s. 4). Vestøl (2008, s. 5) viser til Strand og Kvernbekk (2000, s. 11) sin kritiske gjennomgang av modellen. « ... the model is not theoretically justified» (Strand & Kvernbekk, 2000, s. 11), hvor de teoretiske begrepene brukt i modellen er mangelfullt



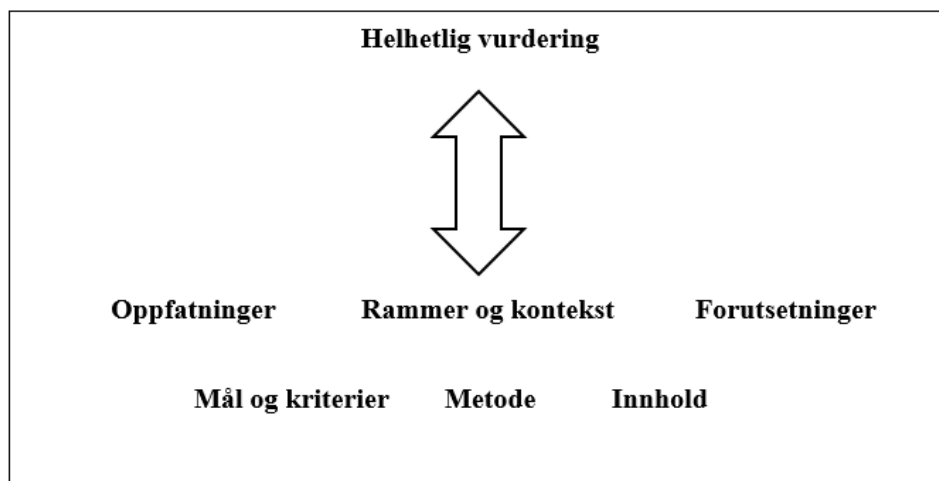
begrunnet. Dermed kan modellen « ... becomes self-sufficient and make teoretical knowledge redundant» (Strand & Kvernbekk, 2000, s. 11). Dette innebærer at det krever en bevissthet å bruke modellen, særlig når struktureringen av teori knyttes til de enkelte elementene.



Figur 7. Relasjonsmodellen knyttet til vurderingens form. Helle (2007, s. 201)

Eksempelet til Helle (2007, s. 201) (figur 7) viser anvendelsen av relasjonsmodellen som et utgangspunkt for å se på vurderingens former. Den kan være til inspirasjon for utvikling av teori om helhetlig vurdering.

Figur 8 viser sammenhengen for HELVYRD-studien og hvordan helhetlig vurdering forholder seg til de ulike didaktiske elementene. I den videre redegjørelsen vil disse elementene bli omtalt som aspekter ved helhetlig vurdering. En kort beskrivelse er gjort hvor min forståelse og tolkning ligger til grunn.



Figur 8. Viser systematiseringen av teori i ulike aspekter om helhetlig vurdering.

Det første aspektet omhandler ulike *oppfatninger* av vurdering mer generelt, samt et mer innledende perspektiv på helhetlig vurdering. *Rammefaktorer og kontekstuelle faktorer* innebærer hva som påvirker vurderingsarbeidet, og hvordan man forholder seg til vurderingsformer og oppgavetyper. Når det gjelder *forutsetninger*, omfatter dette ferdigheter og egenskaper vurdererne har, samt deres rolle og identitet i vurderingsarbeidet. *Mål og kriterier* vil være det som ligger til grunn, og hva det skal vurderes opp mot, samt forventninger og krav som ligger i det utøvende yrket. *Metode* handler om ulike framgangsmåter og strategier hos vurdererne for å gjennomføre vurderingene. Det siste aspektet, *innhold*, er hva vurdererne legger vekt på av faglig kompetanse. De ulike aspektene ved helhetlig vurdering blir ytterligere utdypet i de kommende kapitler. Avslutningsvis ved hvert aspekt om helhetlig vurdering, vil jeg belyse valgt teori opp mot de fire epistemologiske retningene redegjort for i kapittel 3.1.1.

En avsluttende kommentar er at gjennom å presentere teori om helhetlig vurdering med støtte i en didaktisk relasjonsmodell vil ulike perspektiver av yrkesdidaktikk bli berørt. I den sammenheng kan det tenkes at den didaktiske relasjonsmodellen kunne blitt utviklet som en mer yrkesdidaktisk modell, siden den allerede er lite teoriforankret. Men i denne studien blir det for ambisiøst. Jeg forholder meg til de yrkesdidaktiske perspektivene som ble presentert i kapittel 2.3, og anvendes inn i kapittel 6.

### 3.1.3 Oppfatninger om vurdering

Først presentere noen generelle oppfatninger om vurdering, samt en mer innledende fase om helhetlig vurdering.

Vurdering fra et sosiokulturelt ståsted vektlegger hvordan sosiokulturelle faktorer påvirker i en vurderingsprosess. Gipps (1999, s. 384), for eksempel, påpeker hvordan et sosiokulturelt syn på vurdering handler om at vurdering er interaktiv, dynamisk og samarbeidende. Det ses opp imot de formaliserte eksterne tester som kan oppleves å være utenfor læringsprosessen, til at man ønsker å gjøre vurderingen til noe som er integrert i læringen. Det tilsier at man bør se på vurdering i et bredere og dypere perspektiv. Det innebærer å lage systemer for vurdering som kan ivareta ulike sider av læringen, og som gir de lærende anledning til å vise sin kompetanse gjennom ulike vurderingssystemer. Videre utdypes det ved å ha en «interpretivist approach» som viser til viktigheten av å anerkjenne kompleksiteten i forholdene mellom de (lærere og elever) som er involvert i vurderingsprosessen. Det argumenteres også for at de

som skal vurderes blir involvert i vurderingsprosessen og reflekterer over egen læring. Dette er et syn på læring og vurdering som involverer flere faktorer.

Shepard (2000, s. 4) viser i sin artikkel til betydningen av å utvikle vurderingen som skjer i klasserommet og som en del av et system som skal fremme læring, og ha en større betydning enn kun testing av kompetanse. I den sammenheng trekker Tanggaard (2004, s. 73) fram at summativ og formativ vurdering bør ha en større sammenheng i læreprosessen. Utviklingen av vurderingsformer bør legge vekt på læreprosessen (Tanggaard & Elmholdt, 2008, s. 113) da undervisningen organiseres etter vurderingsformer som blir brukt (Eraut, 2004, s. 803). Når det gjelder formativ og summativ vurdering, kan den summative også ses på som formativ ved at tilbakemeldinger kan gi informasjon om behov for læring og undervisning.

Boud og Hawke (2003, s. 5) lanserte en tredje form for vurdering som handler om å promotere livslang læring kalt "sustainable assessment". Boud (2000, s. 151) definisjon er at det er vurdering "... that meets the needs of the present without compromising the ability of students to meet their own future learning needs" (Boud 2000, s.151). Vurdering handler ikke bare om å måle på et bestemt tidspunkt, men også om å påvirke egen læringspraksis. Et av hovedpoengene er at de lærende skal vurdere seg selv og bli involvert i egen vurderingsprosess. Boud og Hawke (2003, s. 5) sier målet er å erstatte eksisterende vurderingsformer med former som fremmer vurdering med et livslangt læringsperspektiv med positive effekter på læring over lang tid. Forholdet mellom vurdering og læring bør ikke være noe misforhold, men å slå fast hva som er den perfekte vurderingsformen er ikke mulig (Eraut, 2004, s. 803).

Black (2000, s. 410) diskuterer «high-stakes» summative tester hvor hensikten er å utvikle testing og vurdering som kan være relevant når det gjelder hvordan og hva man har lært. Det er behov for et bredt spekter av testformer (Black, 2000, s. 413). Shepard (2000, s. 6) rekonseptualiserer vurderingen for å gjøre den mer effektiv gjennom å lansere prinsippene om en sosial konstruktivistisk ramme. Shepard (2000, s.8) uttaler at et bredere spekter av vurderingsmetoder vil gi vurdering en tettere tilknytning til det som har blitt eller skal bli lært. Hun peker på reformer som har gitt mer åpne oppgaver. Disse skal få elever til å tenke mer kritisk, løse mer komplekse oppgaver, og å knytte kunnskapen til reelle kontekster. Hennes bidrag handler om å utvikle en vurderingskultur som skal være noe mer enn bare eksterne tester og som skal brukes mer formativt i klasseromssituasjoner. I den forbindelse tenker jeg at Shepards konseptualisering også vil være nyttig og samtidig kunne påvirke tenkningen

rundt den mer summative og eksterne vurderingen. I denne sammenheng viser Ecclestone (2010, s. 339) til yrkesutdanningen i England, hvor utviklingen av de summative vurderingsformer bygger på kriterie-baserte systemer som kombinerer modeller for kompetanse i arbeidslivet og formalisert utdanning. Det interessante er muligheten for at en ved å nytte det beste fra arbeidslivet og fra skole kan utvikle vurderingssystemer med mer valide og reliable resultater. Gjennom å argumentere for en helhetlig kompetansebasert utdanning uttaler Biemans et al. (2009, s. 269) at flere stakeholders innen VET-feltet forventer et redusert gap mellom yrkesutdanning og nødvendig kompetanse i arbeidsmarkedet. Det presiserer tydelig at tradisjonelle vurderingsmetoder ikke passer til å måle kompetanse og at nye metoder må utvikles. Dette vil kreve en arbeids- og tidskrevende gjennomføring og vil involvere mer enn klasseromseksaminering. Det vil være en kompleks sak å utvikle relevante og pålitelige instrumenter for vurdering av kompetanse.

Når det gjelder helhetlig vurdering uttrykker Dobson (2010, s. 32) at det er spesielt behov for dette innen yrkesfag, siden det er flere kompetanser som skal vurderes i kombinasjon med den faglige kompetansen. Selv når man bryter vurderingen ned i små enheter av kriterier, vil det kreves en helhetsvurdering til slutt. Dobsen viser til flere begreper i debatten om helhetsvurdering som «profesjonelt skjønn», «bruk av intuisjon» og «magefølelse». Kunnskapsdepartementet (2011, s. 25) peker på at all vurdering innebærer bruk av skjønn og utviklingen av profesjonelt skjønn for en lærer handler om kompetanse.

I tilknytning til helhetlig vurdering vil Wolf (1995) sine betraktninger om kompetansebasert vurdering anvendes. Argumentet for en slik vurderingsform ligger i Wolf (1995, s. 128) sin oppfatning om at de vanligste vurderingsformene innen utdanning ikke måler hva de skal måle, hvor det er viktig å tilpasse vurderingsformene til de situasjonen som skal vurderes. Wolf (1995, s. 1) har i den sammenheng redegjort for en definisjon på kompetanse-basert vurdering og som lyder:

Competence-based assessment is a form of assessment that is derived from the specifications of a set of outcomes; that so clearly states both the outcomes- general and specific – that assessors, students and interested third parties can all make reasonably objective judgments with respect to students achievement of nonachievement of these outcomes; and that certifies student progress on the basis of demonstrated achievement of these outcomes. Assessment are not tied to time served in formal educational settings. (Wolf, 1995, s. 1)

Definisjonen utgjør en ramme for den videre drøftingen Wolf har omkring vurdering av kompetanse. Wolf (1995, s. 3) presiserer at kompetanse-basert vurdering er «vocational» i sin form og den har sitt rasjonale med utgangspunkt i profesjonell og arbeidsbasert kompetanse. Kjernen i kompetanse-baserte tilnærminger har utgangspunkt i utdanninger knyttet til profesjon og yrker, hvor konseptet kompetanse kan bli meningsfylt. Kompetansebasert vurdering er videreutviklet fra den kriteriebaserte, men disse ses ikke på som synonyme. Den kriteriebaserte er knyttet til den konvensjonelle skolen. Den kompetansebaserte vurderingen knyttes på sin side mer til det ikke-akademiske, hvor vurderingen utføres i forhold til utøvelsen av yrket. Det prinsipielt viktige i kompetanse-basert vurdering handler om hva som skal vurderes og hva som er oppnådd, og skal formidle en spesifikk informasjon om hva noen har evnen til å gjøre.

Det innledende aspektet om vurdering kan sies å vise et mer overordnet syn, hvor det i et bredere perspektiv diskuteres muligheten for å utvikle vurderingsmetoder som kan tilpasses behovene i en kompetansebasert kontekst. Det vil være nødvendig å ta i bruk flere vurderingsmetoder for å kunne vurdere alle sider ved en kompetanse, hvor både skole og arbeidsliv kan være viktige aktører i prosessen med å lage relevante systemer. Dette kan innebære mer kompleksitet i utviklingen av vurderingssystemer, og vil være en mer arbeids- og tidskrevende form. Dobsons (2010) uttalelse om å benytte helhetlig vurdering ved yrkesfag vil være sentralt og kan knyttes nært opp til Wolfs (1995) redegjørelse om kompetansebasert vurdering. De epistemologiske retningene som kommer til syne i disse teoriene som benyttes i avhandlingen er i størst grad knyttet til det sosiokulturelle, hvor Gipps (1999) kan trekkes frem som en av representantene. Spesielt fremkommer det ved de kritiske blikkene på eksisterende vurderingsformer, og derigjennom ses et behov for å utvikle og tilpasse et bredt spekter av nye vurderingsformer som kan anvendes i ulike kontekster noe flere av forfatterene forfekter. på vurdering gjennom ønsket å ha en tett kontakt med arbeidslivet Eccelstone () og Biemans et al () kan tas til inntekt for et mer konstruktivistisk og instrumentelt syn ved at de ser på praksis i arbeidslivet som det sentrale valideringsgrunnlaget for kompetanseutvikling. En kompetansebasert vurdering som Wolf () er en representant for kan også knyttes til disse retningene.

De teoretiske redegjørelsene ved oppfatninger om vurdering vil bli benyttet mer gjennomgående i de enkelte diskusjonene i kapittel 5, og i den sammenfattende diskusjonen i kapittel 6.

### 3.1.4 Innhold

Helhetlig vurdering vil ha et teoretisk aspekt om faglig innhold.

Wolf (1995, s. 32) viser til de fordeler den teoretiske kunnskapen kan ha, hvor enkelt den kan formidles gjennom en mer standardisert metode, men også hvilken status denne type kunnskap har i utdanningshierarkiet. Dette kan sette press på hvordan kompetanse knyttet til yrkesmessige utførelser blir vurdert, og viser til behovet for å endre dette synet. Det kan få følger for innhold som blir valgt at rammene og konteksten er så avgjørende.

Rauner et al. (2013, s. 23) viser til diskusjoner om innhold i profesjonelt arbeid og læring hvor to studier sammenlignes (Petersen og Rauner: Lindberg). Det kommer fram at begge arbeidene om innhold i profesjonelt arbeid viser en slående likhet. For å utøve yrket viser hovedområdene i de to studiene at verktøy er sentralt, dessuten materialer, metoder og teknikker. Videre er planlegging noe som innebærer å forstå oppgaven og organisere det påfølgende arbeid. Det siste som trekkes fram er knyttet til etikk sett i forhold til eget yrke. Hager, Gonczi og Athanasou (1994, s. 8) mener at yrkesutøvelse innbefatter kunnskaper, forståelse, problemløsning, tekniske ferdigheter, holdninger og etikk, noe som underbygger flere av de ovennevnte hovedområder. Johnson (2008a, s. 29) uttrykker at: «... it is important to consider how assessors integrate and combine different aspects of an holistic performance into a final judgment» (Johnson, 2008, s. 29). Dette er vesentlig med tanke på de innholdskomponenter som kan ha betydning for den enkelte vurderer, og som trekkes inn i den endelige vurderingen. Black (2000, s. 412) viser til hvilken betydning det har at innholdet i testen reflekterer det den skal teste. Noe av det som studentene har lært i skole kan ikke testes ved å bruke eksterne tester. Blant annet kan praktiske ferdigheter ikke kan gjenspeiles i skriftlige tester. Dette får betydning for innholdsdiskusjonen og hvordan vurdererne vil kunne ta hensyn til denne. Eraut (2004, s. 804) trekker fram hvor forskjellig de lærende kan være i sin egen læringsprosess, og det de lærer kan være forskjellig på ulike nivåer. Hvordan kan dette ivaretas og hvordan kan vurderingen bli mer fleksibel? Spørsmålet blir stilt om tidspunktene for gjennomføringen av vurderinger kan bli mer fleksible, og om kriterier for vurderingen bedre kan tilpasses læringsprosessene. I tilknytning til ovennevnte spørsmål trekker Eraut inn begrepet «learning trajectories» (s. 804) som indikerer læringsbanen til den lærende der vurderingsformene tilpasses. Dette er sett i forhold til formative vurderinger hvor argumentet er å få til løsninger som vil kunne gi mer reliable vurderinger underveis. Samtidig kan det bidra til å utvikle gode systemer for summative vurderingssystemer med mulighet for å teste det aktuelle innholdet den lærende har. Hiim (2012, s. 127) tar opp relevans i fag- og

yrkesopplæringen og ulike aspekter som påvirker denne. Betydningen av at det faglige innholdet forankres i arbeidspraksis for elever er viktig, slik at det som læres i skole og på arbeidsplass har en sammenheng. Samtidig kommer det fram at tolkninger av læreplaner og de ulike programfagene (s. 131) gir utfordringer når det gjelder hva det faglige innholdet skal innebære. Hovedkonklusjonen til studien for å oppnå en relevant yrkesutdanning er at det «... må det etableres en struktur hvor skoler og bedrifter samarbeider om utdanningens innhold gjennom hele det faglige løpet» (Hiim 2012, s. 135). En utvikling av helhetlig yrkeskunnskap må skje gjennom en praksisforankring.

Det teoretiske aspektet om innhold framstår som mindre omfattende enn de andre perspektivene. Det kan begrunnes med at innsamlet empiri er direkte knyttet til et konkret yrke med et spesifikt faglig innhold. Men enkelte generelle trekk kan diskuteres. Blant annet gir Rauner et al. (2013) sammenligning av studier innspill om innhold, dessuten Johnson (2008) sin påpekning av hvordan den enkelte vurderer trekker inn sine innholdskomponenter. I den sammenheng vil Erauts «learning trajectories» og den lærendes læring av innhold være et viktig moment i vurderingene. Hiims (2012) har også pekt på at det er viktig med tett kontakt mellom skole og arbeidsliv, for å utvikle det faglige innholdet.

I valget av teori om helhetlig vurdering av faglig innhold fremkommer først og fremst et konstruktivistisk syn som ligger til grunn for både å forstå hva den enkelte innehar (Eraut), men også den tette forbindelsen til arbeidslivet slik Hiim påpeker. Men en forhåndsdefinert forståelse av hvilket faglig innhold som bør være til stede, noe både Johnson og Rauner et al viser til, kan knyttes mer til et disiplinorientert og instrumentelt syn.

### 3.1.5 Rammer og kontekstuelle faktorer

Som en del av helhetlig vurdering vil det teoretiske aspektet om påvirkningen fra de ulike rammene og kontekstuelle faktorene ha betydning. Dette må ses i sammenheng med vurderingssystemenes utforminger og oppgavene som blir brukt. Først kommer her en generell innledende fase om vurderingsformer, og deretter en gjennomgang av mer spesifikke kvalitetskriterier for bruk i utvikling av vurderingssystemer og oppgaver.

Dobsen (2010, s. 34) sier man må spørre om selve vurderingen er god og ikke bare hvordan det går med eleven eller lærling. Hager et al. (1994, s. 8) viser til at det må stilles spørsmål om det finnes en integrert vurdering som kan brukes for å gjenspeile den kompetansen som skal vurderes. Kompetansebaserte vurderingssystemer (Wolf 1995, s. 31) har en nær forbindelse med den situasjonen vurderingen foregår i, den arbeidssituasjon kompetansen skal brukes i.

Imidlertid understrekes det at vurderinger foregår på bruddstykker av en kompetanse (Wolf 1995, s. 42). Dermed blir det vesentlig å lage systemene for vurdering så godt designet som mulig. Jo mer informasjon man får jo bedre grunnlag er det for å gjennomføre en vurdering (Wolf 1995, s. 41). Wolf (1995, s. 114) skriver om kompetansebaserte systemer:

What does matter is that the assessment techniques should be derived from the nature of the competence, and not taken as given. This means that there should also be no automatic presumption in favour of practical tests at the expense of written ones. (Wolf 1995, s. 114).

Jo mer rigide krav systemene er basert på, jo mer kan de hindre en effektiv og høykvalitets vurdering. Wolf uttrykker dette slik:

[...] but no system of assessment and qualifications can, on its own, alter and restructure a labour market, or change the way people behave together. If the system is fundamentally at odds with these it is the assessment system, and the quality of its judgments, which will suffer. (Wolf 1995, s. 125)

Hiim og Hippe (2001, s. 235) uttaler at den avsluttende og formelle vurderingen må ta utgangspunkt i yrkeskunnskap. De poengterer at å utvikle former for vurdering som ivaretar dette er utfordrende. Vurderingsformene bør «... innbefatte utøvd, begrunnet og problematisert yrkesvirksomhet – altså både handling og verbalisert refleksjon» (Hiim og Hippe 2001, s. 236). Vurderingssituasjonene bør knyttes til den reelle yrkesvirksomheten. Ulike former for vurdering som blir nevnt innbefatter både fagprøven og skriftlig eksamen (Hiim og Hippe 2001, s. 235). Skriftlig eksamen bør være oppgaver som har yrkespraksis som utgangspunkt, og som gir rom for refleksjon og begrunnelser knyttet til ulike sider av praksis. Det er en forutsetning at de lærende har hatt anledning til å gjøre relevante erfaringer innen yrket, og at dette kan brukes som referanse når de skal gjennomføre vurderingen. De lærende opplever skriftlig eksamen som meningsfylte hvis det er oppgaver de har hatt erfaringer med. Hiim og Hippe (2001, s. 235; 2009, s. 127) viser til betydningen av å tilpasse vurderingen. Elever og lærlinger må få mulighet til å vise hva de har lært i sin utdanning.

Flere forfattere viser til kvalitetskriterier som bør ligge til grunn ved vurderingssystemer (Baartman, Bastiaens, Kirschner & van Der Vleuten, 2007, s. 15-25; Engh, Dobson & Høihilder, 2007, s. 45-49; Haasler & Erpenbeck, 2008, s. 771; Rauner et al., 2013, s. 56; Wolf, 1995, s. 2). Baartman et al. (2007, s. 2) sier det trengs flere metoder for å vurdere



kompetanse, noe Hager et al. (1994, s. 12) støtter. De uttrykker også behovet for å fylle igjen vurderingsgapet med flere metoder for vurdering. *Competence Assessment Programmes (CAP)* innebærer en kombinasjon av ulike vurderingsmetoder for å få fram klare kvalitetskriterier. Disse programmene er basert på en gjennomgang av ulike synspunkter på vurdering fra flere forfattere. Jeg tar utgangspunkt i de utviklede kriteriene jeg ser som sentrale for min studie, og gir en gjennomgang av disse. Samtidig trekker jeg inn andre ulike forfatteres syn.

Det første kvalitetskriteriet handler om *autentisitet* (Baartman et al., 2007, s. 15). Det innebærer at vurderingen skal reflektere de behov som trengs i en framtidig arbeidsplass. De oppgaver som brukes må underbygge prosessene og strategier som ligger til grunn for det som skal vurderes. Baartman et al. (2007, s. 15) viser til Gulikers, Bastiaens og Kirschner (2004, s. 69) som med sin definisjon av autentisk vurdering uttrykker at studenten må bruke de samme kompetansene i en vurderingssituasjon som det som benyttes i en reell arbeidssituasjon. En autentisk oppgave utfordrer kandidaten med de samme oppgavene som man møter i virkeligheten. Wolf (1995, s. 2) viser til betydningen av å ha en «... decoupling of assessment from particular institutions or learning programmes», noe også Rauner et al. (2013, s. 56) viser til i utviklingen av oppgaver. Han peker her på arbeidslivet og ikke på skoler og utdanningsinstitusjoner, da behovet for oppgaver skal dekke et realistisk problem knyttet til aktuelt yrke. Dette innebærer ikke bare teknisk kompetanse, men også effektivitet, miljø og HMS. Hayton og Wagner (1997, s. 71) viser til bruk av prestasjonsbaserte oppgaver som skal reflektere den reelle arbeidsplassaktiviteten, og støtter at det skal måle mer enn bare kunnskap. Vurderingen skal være multidimensjonal basert på praktiske aktiviteter som kan ende ut i et produkt eller i en prosess. Med case-basert vurderinger viser Segers, Martens og van den Bossche (2008, s. 1752) til hensikten om å få til en dypere læring. Men vurderinger basert på konstruktivistiske prinsipper gir ikke alltid det (s. 1761). Haasler og Erpenbeck (2008, s. 771) tar opp bruk av «*evaluation tasks*» til å vurdere yrkeskompetanse. Styrken ligger i å være virkelighetsnær når det gjelder de aktuelle arbeidsprosesser, og kravene om å gi relevante og realistiske arbeidsoppgaver slik at resultatet av oppgavene gir evidens på deres yrkeskompetanse.

Det andre prinsippet viser til *rettferdighet* (Baartman et al., 2007, s. 16), som innebærer at alle skal ha like muligheter til å få vist sine evner og potensiale. Da blir det viktig at oppgavene er tilpasset de aktuelle studentene. En test skal dekke alle kunnskaper og ferdigheter og de strategier som skal til for å svare bra. Vurderingen skal reflektere alt dette.

Engh et al. (2007, s. 34) viser til rettferdighet som et kriterium, og dette handler om å ha en vurderingsform som sikrer at alle kan få vist hva de kan. Rauner et al. (2013, s. 56) viser til differensiering av oppgaver med det målet å tilpasse de ulike nivåene (noviser, avanser nybegynner, kompetent utøver og profesjonell utøver). Man skal utfordre kandidatene til å utføre oppgavene profesjonelt. Hayton og Wagner (1997, s. 71) mener at det er behov for å gå fra enkle til komplekse aktiviteter i vurderingen.

Neste kriterium er *meningsfylde* (Baartman et al., 2007, s. 16). Vurderingen skal ha en signifikant verdi for både lærer og studenter. Ved en meningsfull vurdering må studentene oppleve at de får jobbe med utfordringer som gir mening og som har en link til deres personlige interesser. Engh et al. (2007, s. 45) viser til tilpasset hensikt hvor vurderingen må treffe i forhold til det formålet, og at vurderingsformen er «suited to the task» (Dobson 2010, s. 35). Knyttet til prestasjonsbaserte oppgaver uttrykker Lai (2011, s. 1) at vurdering relatert til formål er mer motiverende, og skal kunne måle evnen dypere for de lærende. Hager et al. (1994, s. 13) støtter dette i sammenheng med kompetansebasert vurdering, der prestasjoner står i fokus. Segers et al. (2008, s. 1761) uttrykker at dypere læring vil oppstå gjennom casebaserte vurderinger når de lærende har oppfattet kravene. I et konstruktivistisk perspektiv vil det kreve aktiv deltakelse og engasjement hos de lærende (Lounsbury & Pittenger, 2011, s. 1). Struyven, Dochy, Janssens, Schelfhout og Gielen (2006, s. 219) viser til den utfordring (komplekst og problematisk) det ligger i det å vurdere en lærende. Det trengs en triangulering av vurderingsmetoder for å få til en god vurdering.

Det neste kriteriet, *direkthet* (Baartman et al., 2007, s. 16), handler om i hvilken grad lærere og studenter kan tolke vurderingsresultatet uten å gå fra teori til praksis. En direkte vurdering av handlingen vil kunne styrke validiteten i vurderingen. Hassler og Erpenbeck (2008, s. 769) uttrykker at kompetanse kan bare bli vurdert i den konteksten man skal løse de aktuelle arbeidsoppgaver, eller i et miljø som kan simulere de aktuelle arbeidsoppgavene. Rauner et al. (2013, s. 56) og Hayton og Wagner (1997, s. 71) ønsker åpenhet for at flere løsninger kan forekomme og at noe ikke er riktig eller galt. Dette handler om å se situasjonen slik kandidaten har utformet den.

Det neste kriteriet er *transparent* (Baartman et al., 2007, s. 17), som betyr at oppgavene må være klare og forståelige for de involverte parter. Også kriteriene skal være klare, hvem som skal vurdere bør være kjent, dessuten hva hensikten med vurderingen er. Forventninger bør være tydelige for de lærende, og for de som skal vurdere. En indikasjon på transparenss kan

handle om at studentene kan bedømme seg selv like nøyaktig som en vurderer vil gjøre. Engh et al. (2007, s. 48) påpeker betydningen av gjennomsiktighet slik at de involverte parter i vurderingsarbeidet får anledning til å utvikle et tolkningsfelleskap. Wolf (1995, s. 2) uttrykker at “The belief that these can and should be specified to the point where they are clear and «transparent»”. Målet er at de involverte partene er klar over hva som skal vurderes og hva som er oppnådd (Wolf 1995 s. 2). Wolf (1995, s. 2) viser videre til at «The emphasis on outcomes» og spesielt «multiple outcomes» er vesentlig når det knyttes til kompetansebasert utdanning.

Kriteriet *reproducibility of decisions* (Baartman et al., 2007, s. 17) er en term som brukes for å vise at det trengs et mangfold av vurderingsmetoder med flere vurderere i ulike sammenhenger. Dette handler om generaliserbarhet og formidling av resultater til en bredere krets av studenter. Det handler ikke bare om en vurderers vurdering i en bestemt situasjon, men at vurdereren skal kunne trekke en mer generell konklusjon om studentens kompetanse som inkluderer mange vurderingsformer. Engh et al. (2007, s. 48) viser til gyldighet ved vurderingen og de formene man bruker for å avdekke kompetanse. Noen former kan vurdere alt for lite i forhold til det kandidaten har levert. Rauner et al. (2013, s. 3) viser til hva man kan få til ved bruk av tester med penn og papir test. Han uttrykker at KOMET-modellen er en test på konseptuelt nivå, og ikke som vurderer praktiske ferdigheter. I motsetning viser Hayton og Wagner (1997, s. 69) til at penn- og papir-tester ikke er tilstrekkelige for å utvikle en kompetanse. Også Hager et al. (1994, s. 7) understreker at dette kan bli for snevert.

Det siste kriteriet handler om *sammenlignbarhet* (Baartman et al., 2007, s. 18), at vurderingen i mest mulig grad skal gjøres lik for alle studentene, og at det blir en konsistent skåre på resultatene. Siden det er mindre standardiserte vurderingssystemer i CAP, vil det i større grad baseres på den menneskelige faktoren i bedømmelsene. Hayton og Wagner (1997, s. 71) underbygger dette med å peke på hvor viktig det er at vurderingen gjøres av mennesker. Sammenligningen kan bli nådd når vurderingen er basert på et sett med oppgaver som ikke bør være identiske, og som er konsistente når det gjelder de viktigste funksjonene. Lai (2011, s. 20) presiserer at prestasjonsbaserte oppgaver med mulighet for å kunne sammenligne disse med hverandre kan være et viktig aspekt. Pålitelighet er viktig, at de prøver og oppgaver som benyttes er likt hver gang (Engh et al., 2007, s. 47).

Oppsummert vil de utvalgte kvalitetskriteriene til Bartmaan (2007) være sentrale i diskusjonen om rammer og kontekstuelle faktorer ved helhetlig vurdering.

Vurderingssystemene inkludert oppgaver utgjør sentrale rammer og kontekstuelle faktorer i HELVYRD-studien. Spørsmålet om hvordan dette påvirker vurderernes helhetlige vurderinger er vesentlig å få belyst og diskutert, ikke minst med tanke på handlingsrommet i utøvelsen av vurderingene. Gjennomgangen av Wolf, Hiim og Hippe og Bartmann et al. (2001) sine kriterier viser at det ligger både et konstruktivistisk og sosiokulturelt syn til grunn. En nær tilknytning til arbeidslivet hvor vurderingsformene skal tilpasses ulike kontekster for å ivareta det som skal vurderes. Det påpekes også her nødvendigheten av å ha flere vurderingssystemer noe som er i tråd med det som ble tatt opp under oppfatninger om vurdering.

### 3.1.6 Mål og kriterier

Når det gjelder helhetlig vurdering og det teoretiske aspektet om mål og kriterier, handler dette om hva det skal vurderes opp mot og hvordan det kan gjøres.

Wolf (2001, s. 2) har i sine diskusjoner vist til forholdet mellom det å gjøre kompetansebasert vurdering sett opp imot nasjonale kriterier (NVQ-systemet). Hvis ikke ferdig skrevne definisjoner av kompetanse kan gi tydelig nok forklaringer på hva som forventes, er alternativet å støtte seg til en forhåndsforståelse og konsensus mellom vurdererne. Wolf argumenterer med at standarder ikke kan beskrives så konsist at alle kan bruke dem i utviklingen av en reliabel og konsistent vurdering. Det blir påstått at kompetansebasert vurdering er lett å gjøre for vurdererne fordi kriteriene er så klart spesifiserte, og at vurderingen dermed kan gjøres på en mekanisert måte. Dette stemmer ikke (Wolf, 2001, s. 8). Vurderinger som gjøres på ulike arbeidsplasser og kontekster varierer, og vurderingsprosessene er komplekse. Wolf mener at kompetansebasert vurdering blir en utfordring når det er for ambisiøst fokus på å lage en nasjonal standard og sentral kontroll over innholdet (s. 10). Dette begrunnes med at man viser for lite innsikt i vurdering av enkeltmennesker, og at det kreves at man forstår hvordan grundige analyser av arbeidsoppgaver og prosesser skal gjennomføres (Wolf, 2001, s. 10). Wolf retter kritikk mot holdningen om å beskrive et omfattende arbeidsdomene (detaljert eller bredt). Jo mer beskrivelsene utformes jo mer begrenset blir domene i seg selv, og dermed vil det bli mindre transparent (Wolf, 1995, s. 55). Wolf (1995, s. 55) uttrykker at «The attempt to map out free-standing content and standards leads, again and again, to a never-ending spiral of specification» (Wolf, 1995, s. 55). Kriterier må tilpasses og utvikles fortløpende i samsvar med den konteksten der vurderingen skal skje. Samtidig kan en kartlegging av spesifikasjonene i et fagdomene gjøre det mer snevert. Det er meget krevende å skrive ut all den informasjonen som trengs. Det vil dessuten lett oppstå tvetydighet (Wolf 1995, s. 56). For stor grad av spesifisering med kriterier

og fagdomener vil det kunne være selvødeleggende for en kompetansebasert vurderingen (s. 60). Samtidig vil det i en verden hvor kompetansebasert vurdering benyttes allerede eksistere «standarder» på kompetanse. «Rather, they articulate and clarify them to professionals who can understand them because of their prior knowledge and implicit understanding of what “competence” in their own context means” (Wolf, 1995, s. 65). Ved en felles forståelse av faget hos ekspertene kan den endeløse spiralen av kriterier gjøres overflødig. Wolf uttrykker følgende om bruk av kriterier og standarder:

Nonetheless, we must seriously question whether it is likely even in principle that a combination of definitions and prior consensus will produce any very uniform behavior. We must also wonder whether the assumptions of pre-existing standards and shared understanding is reasonable at all. (Wolf, 1995, s. 65).

Wolf kan tolkes slik at man må være kritisk til om det er mulig og hensiktsmessig å etterstrebe en utforming av kriterier som vil frambringe en «uniformering» og tydelig delt forståelse hos vurdererne ved kompetansebasert vurdering. Som nevnt tidligere kan det være mer hemmende enn fremmende ved selve vurderingsprosessen.

Dobsen (2010, s. 31) viser til at det kreves et fellesskap for å tolke kriterier og bruken av disse. Alle bør ikke være enige, men det er et poeng at alle skal kjenne til hverandres synspunkter på kriteriene. Ofte kan utviklingen av kriterier og standarder utformes for å tilfredsstillere det valgte vurderingsinstrumentet for elever og lærlinger. Internasjonalt er det flere land som utvikler detaljerte sett med kriterier som gir en gevinst gjennom en mer tydelig forbindelse mellom kompetansemål, læring og vurdering, hvor vurdererne får en mer felles, helhetlig og rettferdig vurderingspraksis (Dobson, 2010, s. 32). Ved bruk av kriterier kan man begynne analytisk og krysse av for hvert enkelt kriterium, men dette må ses i en holistisk sammensetting for å sikre en samlet vurdering til slutt (Dobson, 2010, s. 32). I motsatt tilfelle kan besvarelser først leses gjennom for å skaffe seg et helhetsinntrykk, hvor det i etterkant kan finnes bekreftende eller avkreftende kriterier. Internasjonalt kan det vises til en pendelbevegelse hvor det på den ene siden er det analytiske med vektlegging av kriterier, mens det på den andre siden legges mer vekt på det holistiske (s. 33).

Colley og Jarvis (2007, s. 299) redegjør for kriterier ved summativ kompetansebasert vurdering handler om det formelle, som ender ut i en vurdering, og det uformelle som har læring som mål, og ikke ender opp i en endelig vurdering. De uformelle sidene ved vurdering bidrar stort til den formelle vurderingen, at de lærende når sine mål, noe som delvis skyldes at

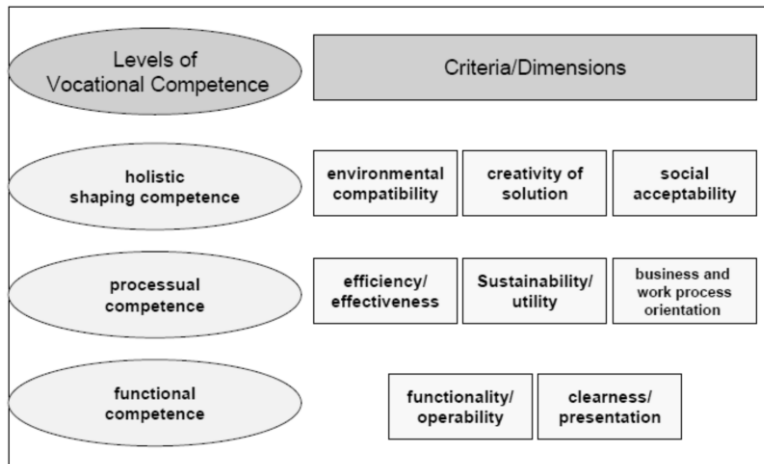
de ressurser som trengs blir supplert. Det presiseres at de uformelle sidene ved vurderingen ikke skal skape grunn til uro om kvaliteten på vurderingssystemene, men heller anses som en styrke og dempe vekten på formalisering. Dette er et positivt bidrag som kan sikre en god vurdering av de lærende (Colley og Jarvis, 2007, s. 312).

Hager (gjengitt i Jones, 1999, s. 154) påpeker at hvis det som skal læres blir for konsist, vil det begrense muligheten for å gjøre en unik og helhetlig vurdering. Da vil mangfoldet i den lærendes kompetanse ikke komme til uttrykk. Et vesentlig element er at det ikke lar seg gjøre å beskrive kriterier fullt ut i forhold til mangfoldige og komplekse kompetanser som kan brukes som en sjekklister. Det å lære og vurdere praktisk kunnskap involverer blant annet taus kunnskap og bedømmelser som er spesifikke for den aktuelle konteksten (Jones, 1999, s. 154). Videre peker Eraut (2004, s. 804) på at sider ved en kompetanse knyttet til utøvelsen av et profesjonelt yrke ikke vil bli tatt med i en vurderingssituasjon. Det kan handle om holdninger, samarbeid, kulturelle faktorer, forpliktelse og selvvurdering, vesentlig elementer for utøvelse av yrker. En svært viktig presisering fra Eraut (2004, s. 804) er at det ikke må ende i en vurdering hvor kompetansen som vurderes vil bli delt opp i små enheter av kunnskaper og ferdigheter som blir vurdert hver for seg. Det er viktig å holde på helheten for å gjenkjenne den komplekse profesjonelle handlingen. Hager et al. (1994, s. 5) støtter uttalelsen om at en økt atomisering av en kompetanse kan i utgangspunktet medføre vanskeligheter for å få til en valid helhetlig vurdering. Samtidig uttaler Hoy-Mack (2005, s. 86) at det kan være en nødvendighet å bryte ned de enkelte kompetanser til atomiserte enheter, for så å se nærmere på samspeillet mellom disse, og derigjennom samle de ulike fragmentene til en helhet.

I den norske konteksten har G. S. Haaland et al. (2009, s. 230) vist til behovet for å etablere gode rutiner for underveis- og avsluttende vurdering, og uttaler at vurderingskriterier skal være med. De er opptatt av spørsmål knyttet til utarbeiding av vurderingskriterier og de utfordringer dette kan føre med seg ved en fagprøve. Hvilke krav på kvalitet stilles til arbeidet på prøven i bedriften der den avholdes? Hvilke kriterier legger den enkelte prøvenemndsmedlem vekt på i vurderingen av god fagutøvelse?

Rauner et al. (2013, s. 47) med sin KOMET- modell (hovedmodell presentert i kapittel 2.2) har utviklet noen dimensjoner som de mener ligger til grunn for å løse alle typer arbeidsoppgaver. Dimensjonene skal gjenspeile og omfavne kompleksiteten ved det mest karakteristiske i enhver yrkesoppgave, og uttrykke hva som forventes av en fagperson

kompetanse for å utføre de ulike arbeidsoppgavene. Dimensjonene er hverken statiske eller endelige, de fungerer sammen i en helhet, og forståelsen av hver dimensjon må sees i forhold til de ulike kontekstene. Figur 9 viser en modell med de åtte dimensjonene (høyre del) som igjen er delt inn i de tre nivåene fra kravdimensjonene (venstre del og uten nominell).



Figur 9. Kompetanse nivå og dimensjoner/kriterier KOMET-modell (Rauner, 2013, s. 48)

Dette er et eksempel på hvordan standarder og kriterier kan utvikles, og disse vil utgjøre grunnlaget for den vurderingen som skal gjøres av arbeidet som er gjort.

Avslutningsvis i dette kapittelet vil det bli gitt noen betraktninger om karakterer som en vesentlig del av summativ vurdering, i den sammenheng knyttet til diskusjonen om kriterier. Hiim og Hippe (2001, s. 243-244) uttrykker at den flerdimensjonale og helhetlige yrkeskunnskapen er kvalitativ i sin karakter. De stiller spørsmålet om det er hensiktsmessig å måle og gradere den i enkelte karakterer. Ved bruk av karakterer vil det være behov for målbare kriterier, noe Hiim og Hippe stiller spørsmål ved. De mener dette kan være problematisk siden yrkeskunnskapen kan bli forvrengt eller overforenklet. Videre stiller de spørsmålet om karaktersetting kan godkjenne mindre egnede yrkesutøvere. Hva man legger inn i de enkelte karakterene er av betydning, og man skal være klar over at de som skal karaktersettes vil oppleve dette på ulike måter. Man kan bli skuffet og føle at vurderingen ikke er god nok i forhold til ens egen opplevelse. Vurdererne er i en maktposisjon og må være seg dette bevisst. Men karaktersetting på summative tester er et samfunnsoppdrag for å informere ulike parter. G. S. Haaland et al. (2009, s. 226) viser til nødvendigheten av å ha mest mulig klar forståelse for hva som ligger i de ulike karakterene (bestått meget godt/bestått/ikke bestått) til fagprøven. Det er viktig å legge arbeid i det å gå opp grenseganger for de ulike karakterene, noe som kan uttrykkes veldig forskjellig fra fag til fag. Williams og Bateman (2003, s. 33) viser til at det er viktig å være «fair» i sin bedømmelse, og

å ta hensyn til kandidaten som skal karaktersettes. Samtidig påpekes det at «stakeholders» innenfor VET forventer et brudd med dette prinsippet, da den karakteren som settes skal gi en trygghet på at denne kandidaten er god nok for høyere utdanning eller arbeidslivet. Men ingen kandidat skal ha noen ulemper i vurderings- og karakter prosessen. Johnson (2008b, s. 182) viser til at ved karaktersetting innenfor kompetansebasert vurdering er to vesentlige faktorer framtrepende. En tydelig verdi for kvalifisering av kandidaten og at vurderingen er valid. Begge faktorene må være til stede, ellers vil vurderingen bli undergravet. Williams og Bateman (2003, s. 34) viser spesielt til betydningen av validiteten og reliabiliteten, og dermed rettferdighet når det gjelder karakterer ved kompetansebasert vurdering. De mener også at et er behov for å øke en felles forståelse og konsistens i tolkningen i karaktersetting ved vurdering. Allen (2005, s. 218) viser til betydningen av å kommunisere den rette karakteren, at den er mest mulig presis i forhold til de vurderinger som er gjort, og at den gjenspeiler den lærendes kompetanse. Spørsmålet om et helhetlig (multidimensjonal) blikk på en lærendes kompetanse gjelder også om en karakter kan gjenspeile dette i en «merged judgment» (Allen, 2005, s. 220). I studien til Williams og Bateman (2003, s. 27) var meningen delt om subjektivitetens betydning ved karaktersetting, men de fleste mente den ikke hadde stor betydning. Det kom tydelig fram at karaktersetting gjorde vurderingen mer kompleks og det var viktig med gode strategier for validering.

Som en oppsummering av helhetlig vurdering knyttet til det teoretiske aspektet om mål og kriterier framstår det sentralt at man ikke lager standarder og kriterier for omfattende. Både Wolf (1995, 2001) og Hager i (Jones 1999) viser til det «umulige» i å skrive kriterier fullt ut for å dekke et helt arbeidsdomene. Wolf påpeker at det er viktig å tilpasse kriteriene til den aktuelle konteksten. Samtidig viser Dobson (2010) og Sund Haaland (2009) til behovet for å utvikle vurderingskriterier, hvor Dobson spesielt mener at det er behov for å jobbe med disse kriteriene i et fellesskap for å skape forståelse. Det vises til en pendelbevegelse som går fra analytisk til holistisk. Colley&Jarvis (2007) peker på viktigheten av den uformelle siden, og betydningen det kan få for den formelle vurderingen. Både Eraut (2004), Dobson (2010), Wolf (1995, 2001) og Hager et al. (1994) understreker hvor viktig det er at det som skal vurderes ikke må bli for atomisert, men ses på som en helhet. Samtidig viser Hoy-Mack (2005) til betydning av å se samspillet mellom de atomiserte elementene, før det til slutt ender i en helhet. Rauner et al. (2013) sin KOMET-modell viser en mer strukturert måte å utvikle kriterier på for å ivareta en helhet av kompetansen. Når det gjelder karakterer, viser Hiim og Hippe (2001) til utfordringene med å måle og gradere en kompleks yrkeskompetanse ved



hjelp av karakteruttrykk. Man skal også være seg bevisst sin rolle og være fair i sin karaktersetting (Williams & Bateman, 2003). Johnson (2008) peker på betydningen av verdi og validitet, og Williams&Bateman (2003) underbygger dette med fokus på reliabilitet og validitet ved kompetansebasert vurdering. Betydningen av vurderernes subjektivitet var ikke framtredd ved karaktergivning. Allen (2005) viser til å kommunisere den rette karakteren, og om det også er mulig i en «merged judgment».

Til grunn for den utvalgte teorien om mål og kriterier ligger det både et konstruktivistisk og sosiokulturelt syn. Samtidig fremkommer det nyanser noe pendelbevegelsen illustrerer. I større grad fremkommer et mer instrumentelt orientert syn tilknyttet Rauner et al (2013) sin redegjørelse, som kan begrunnes i ønsket om å balansere det å ivareta alle sider av en yrkeskompetanse gjennom forhåndsbeskrevet ramme, samtidig ha en viss grad av fleksibilitet og tilpasning. Synet på karakterer kan forstås ut fra en generell humanistisk tradisjon som ofte kan kobles til et sosiokulturelt syn på kunnskap samt et kritisk blikk på kontekstfri anvendelse innenfor et yrkesfaglig felt.

### 3.1.7 Metoder

Helhetlig vurdering med fokus på det teoretiske aspektet om metoder knyttet til vurderingsprosessene handler om et overordnet blikk på framgangsmåter og prosesser i vurderingsarbeid.

Innledningsvis viste jeg til Dobsen (2010, s. 32) som uttrykte at i yrkesfag har man særlig behov for helhetlige vurderinger av både elevers og lærlingers arbeid. Selv om de ulike delene av en tenkt kompetanse brytes ned i mindre enheter, vil det til slutt være ønskelig å gjøre en helhetsvurdering. Dobsen viser til Sadler (2009, s. 7) som uttaler:

A work which the teacher would rate as 'brilliant' overall may not be outstanding on all the preset criteria. The whole actually amounts to more than the sum of its parts. Conversely, a work the teacher would rate as mediocre may come out extremely well on the separate criteria. For it, the whole is less than the sum of its parts. This type of mismatch is not confined to educational contexts. (Sadler, 2009, s. 7)

Selv med god eller dårlig score på kriteriene vil ikke den helhetlige vurderingen automatisk følge etter. Sadler (2009, s. 14) påpeker at det positive ved helhetlig vurdering er åpenhet rundt det å innlemme kriterier, og at den ikke er utformet som en ferdig liste. Man bør unngå å bruke en forhåndsdefinert liste med kriterier på en mekanisert måte. Både lærere og de

lærende burde bli «... inducted into more open ways of making grading decisions and justifying them» (Sadler 2009, s. 14).

Wolf (1995, s. 69) viser til holistiske vurderinger av komplekse arbeidsoppgaver hvor kriteriene blir tatt fram på slutten for å finne ut om det er noe som ikke har blitt berørt i besvarelsen. I vurderingsprosessen er det mange komplekse bedømmelser som skal gjøres, og hovedgrunnen til bedømmelsene ligger mer opp til vurdererens egen standard enn til de satte kriteriene. Poenget til Wolf er at vurdererne kan vurdere i henhold til en akseptabel standard, men det er viktig å være klar over at prosessen er kompleks, trinnvis og ikke minst forutinntatt (Wolf 1995, s. 70). Slik må det være siden den enkelte vurderer observerer en handling ulikt, og kan som nevnt tidligere ikke bare bli plassert mekanisk i en liste med kriterier. I en kompetanse-basert vurdering vil vurdererne alltid forholde seg til et stort antall kriterier som til sammen utgjør en kompetanse (Wolf 1995, s. 70). Dermed vil spørsmålet om «compensation and aggregation» øke betraktelig. Selv med konkrete rammer og standarder som skal følges vil vurdererne være i den situasjonen at de må gjøre en kompleks bedømmelse av tilstrekkeligheten til utøvelsen, noe som ikke kan gjøres på en mekanisk måte.

Hoy-Mack (2005, s. 86) viser til helhetlig og atomisert vurdering, og forholdet mellom den som blir vurdert og den som vurderer. I en arbeidsbasert vurdering vil et sentralt aspekt være at den som gjennomfører vurderingen fortløpende analyserer den lærende, og knytter dette til de ulike kompetansene. Det vil være et samspill mellom det helhetlige og det analytiske i vurderingsarbeidet, noe i likhet med det Dobson (2010, s. 33) viste til som en pendelbevegelse. Et viktig moment i dette arbeidet er at den som vurderer ikke forteller den lærende om de ulike atomiserte kompetansene underveis, men gjør notater for ikke å forstyrre og forsinke utøvelsen den lærende gjennomfører. De atomiserte/modulære kompetansene blir integrert i den naturlige arbeidsplassbaserte utøvelsen.

Når Dobson omtaler nytten av helhetsvurdering, viser han til Eisner (1991, s. 63) som uttrykker med sitt begrep «connoisseurship» at «The ability to make fine-grained discriminations among complex and subtle qualities is an instance of what I have called connoisseurship» (Eisner, 1991, s. 63). Begrepet *kompleks* vil være sentralt i forhold til de oppgaver som skal vurderes innen yrkesfag. Eisner uttaler at «Connoisseurship is the art of appreciation» (Eisner 1991, s. 63). Det handler om å ha et blick på både prosess og produkt i vurderingen noe Hager et al. (1994, s. 13) også presiserer er vesentlig i forhold til vurdering av prosesser. Den som vurderer skal forholde seg til et bredt spekter av kilder og sanser.

Eisner (1991, s. 80-81) stiller spørsmål ved praktisk vurderingsarbeid og hva vi bedømmer, hva som inkluderes og samtidig ekskluderes både i uformelle og formelle vurderingssekvenser. Dobson (2010, s. 33) mener at det som Eisner knytter til begrepet «connoisseurship» passer godt i sammenheng med vurdering av yrkesfagelever, hvor blant annet krav om tverrfaglighet er tydelige. Det bør, uavhengig av fag, være en vid dokumentasjonsbasis når det gjelder alle former for vurdering. Kriterier kan bidra til å snevre innover mens en helhetsvurdering vil kunne åpne opp og være mer inkluderende (Dobson 2010, s. 33).

O'Grady og Alwis (2012, s. 200) peker på at det å bruke flere vurderingsmetoder kan gjøre bedømmingsoppgaven mer kompleks fordi grunnlaget for vurdering kan gi et variert bilde. De uttrykker at « ... assessing understanding is a complex process that is not easily represented and interpret, required multiple points of comparison and corroboration» (O'Grady og Alwis, 2012, s. 201). Ved å bruke bare en metode kan vurderingen oppleves enkel, men hva man får ut av en enkeltstående test vil gjerne være begrenset. Styrken i helhetlig vurdering kan ligge i at man benytter flere metoder for vurderinger for å sikre et større grunnlag for bedømmingen. Det vil kunne gi et større grunnlag for å få innsikt i den lærendes ståsted, og dermed gjøre det sannsynlig at bedømmingen passer til de kriterier som er satt (O'Grady og Alwis, 2012, s. 203).

Skjønn vil være en del av de ulike metoder som benyttes. Grimen og Molander (2008, s. 188) viser til ulike prinsipper som vil virke inn. Det første er *likebehandlingsprinsippet* (s. 188), som handler om komparativ konsistens i de bedømmelser som gjennomføres, og at dette kan utfordres. Forfatterne viser til eksempler på vurdering av eksamen innen fag hvor det ikke er noen fasit. Sensorene er som oftest enige, men det kan også være store forskjeller. Skjønn kan true likebehandling og «... det kan true disse konsistensbetingelsene selv om det blir utøvd så samvittighetsfullt og grundig som vel mulig» (Grimen og Molander, 2008, s. 189). Det andre prinsippet er *individualiseringsprinsippet* (s. 189) som innebærer at man opererer innenfor et normativt rammeverk og at den enkelte får en individuell oppfølging. Selv om situasjoner kan se like ut, kan en skjønnsmessig vurdering innebære ulike handlinger. Det legitime i slike kontekster handler om det spesielle eller unike ved den aktuelle persons situasjon eller behov. Ved skjønnsutøvelse må det forventes en fornuftig uenighet og vil inneholde kilder til ulike konklusjoner (s. 192). *Førstepersonserfaring* (s. 193) handler om at det er forskjell på personer som utøver skjønn. Den enkeltes erfaringsgrunnlag kan være forskjellig basert på ulike faktorer. Når personer erstatter hverandre, kan de derfor ikke forvente at det skal bli

nøyaktig det samme resultatet, men at man kan få noe liknende. Blant annet vil livserfaringer i et mangfoldig samfunn prege det enkelte menneske. Dette kan gi ulike vurderinger i komplekse saker. Grimen og Molander (2008, s. 194) viser til spenningen mellom skjønn og komparative konsistenskrav. De peker på at skjønn som resonnering fører til utfordringer når det gjelder komparative konsistenskrav som er vanskelige å tilfredsstille.

Som en del av helhetlig vurdering viser Vaughan (1991, s. 121) i sin oppsummering på at flere av de som vurderte gjorde komparative sammenligninger med tidligere vurderinger, og at dette var vanlig. Grimen&Molander (2008, s. 193) mener at når avgjørelser i lignende tilfeller er med på å bestemme utfallet, er det viktig å være klar over at de tidligere sakene ofte vil inneholde normativitet. Lokale tradisjoner og kulturer vil påvirke i slike beskrivelser og vil være en kilde til spenninger mellom skjønnsutøvelse og komparative konsistenskrav. Avslutningsvis trekker Hager et al. (1994, s. 6) fram at validitet knyttet til kompetansebasert vurdering handler om å måle det som skal måles. De påstår at en for stor grad av atomisering ikke gir en valid vurdering av den aktuelle kompetansen. De viser til at den profesjonelle vurderer vil spille en sentral rolle, og at både pålitelighet og validitet ikke bør synke av den grunn Hager et al. (1994, s. 14).

I oppsummering om helhetlig vurdering og det teoretiske aspektet knyttet til metoder viser Dobson (2010) til betydningen av helhetlig vurdering innen yrkesfag. Sadler (2009) og Eisner (1991) synspunkter på helhetlig vurdering viser til kompleksiteten, og at flere elementer spiller inn i den endelige vurderingen. «Connoisseurship» viser til betydningen av et bredt spekter med kilder og sanser i vurderingsprosessen. Kriteriene kan ha en avgjørende betydning i helhetlig vurdering, ikke minst når det gjelder å unngå en mekanisk innstilling til selve vurderingsprosessen. Dobson uttrykker at en helhetsvurdering vil åpne utover hvor kriterier kan snevre innover. Dette støtter også Wolf (1995), som presiserer at en vurdering ligger nærmere opp til vurdererens standard enn de utviklede kriteriene. Å bruke flere vurderinger er vesentlig, og O'Grady og Alwis (2012) påstår at flere av kriteriene vil passe. Det er viktig med samspillet mellom de som vurderer og de som blir vurdert. Dette utgjør en pendel mellom det helhetlige og det atomiserte (Hoy-Mack, 2005). Grimen og Molander (2008) har sett på ulike sider ved skjønn, hvor tre ulike prinsipper er vesentlig å ha med seg. Det er likhetsprinsippet, individualiseringsprinsippet og førstepersonserfaring. Vaughan (1991) viser til at sammenligninger forekommer hos vurdererne. Grimen og Molander (2008) peker på at den normativiteten i tidligere eksempler er avhengig av kultur og tradisjoner, og har innvirkning på bedømmelsene. Så ved perspektivet om metode handler dette om å se ulike

former som benyttes av vurdererne, og hvordan disse påvirker selve vurderingsprosessen. Skjønn vil stå sentralt, og at vurdereren bruker «fine-grained» beslutninger i sine vurderinger, og samtidig opprettholder et holistisk blikk på det som skal vurderes. Hager et al. (1994) støtter det og viser til betydningen av den profesjonelle vurderer. Ut fra et epistemologisk syn viser den utvalgte teorien om helhetlig vurdering ved metoder at det til grunn ligger et konstruktivistisk syn, spesielt gjennom Wolf med støtte fra Dobson, O'Grady og Alwis og Hoy-Mack. Eisner om «connoisseurship» og Grimen og Molander sitt skjønn ligger tydeligere over mot en sosiokulturell forståelse av kunnskap.

### 3.1.8 Forutsetninger

Som en del av helhetlig vurdering vil det teoretiske aspektet om forutsetninger innebære at det legges vekt på vurderernes grunnlag og identitet for de valg som gjøres i vurderingsarbeid, og hvordan det kan innvirke på resultatene til de lærende.

I Kunnskapsdepartementet (2006, s. 79) blir vurderingskultur omtalt som for svak. Utdanningsdirektoratet (2009, s. 7) om et helhetlig system for vurdering viser til definisjonen av vurderingskultur: «Med vurderingskultur forstår vi hvordan skolen oppfatter og praktiserer vurdering, hva som vurderes, hvem vurderingen får konsekvenser for og hvordan skolen bruker resultatene.» (Utdanningsdirektoratet, 2009, s. 7). Som en forlengelse uttrykkes det at *vurderingskompetanse* handler om bruk av ulike metoder og vurderingsformer samt didaktiske overveielser. Vurderingen som skal gjøres må stå i sammenheng med de målene som er satt for det som skal læres.

Wolf (1995, s. 67) viser til vurdereres oppførsel og den aktive rollen deres egne konsepter og tolkninger spiller inn i vurderingen. Vurdererne tar ikke den enkelte kandidat på en mekanisk måte og sammenligner den opp imot vurderingsinstruksjonene. Tvert i mot uttrykker Wolf at vurdererne opererer med en internalisert og et holistisk sett med konsepter om hva de mener en vurdering skal vise. Det tas hensyn til den konteksten som den konkrete handlingen skjer i, samt at det refereres til andre evidenser relatert til kandidaten for å avgjøre hva de egentlig mente i sine handlinger og avgjørelser. Vurdererne vil ta i betraktning om oppgaven eller spørsmålene var spesielt vanskelig ved vurdering av kandidatens respons.

Kompleksiteten i vurderingssituasjonen og variasjoner i konteksten der vil påvirke. Vurdererne må hele tiden ta viktige avgjørelser. De må avgjøre hvordan de skal ta hensyn til konteksten ved en bedømming, og om det observerte passer i forhold til kriteriene. De opererer i en kompenserende prosedyre som krever en internalisert og holistisk modell, ikke

et enkelt sett med atomiserte enkelt elementer. Wolf viser til at vurdererne i en vurderingssituasjon «... compensate, make allowances, interpret, explain away» (Wolf 1995, s. 71). Jo mer erfaren vurdererne er, jo mer opererer de i et fellesskap hvor de har internalisert en modell for kompetanse, og jo mer aktiv vil deres aggregerende vurdering bli. Som et eksempel vises det til et fagfelt hvor uerfarne og erfarne studenter skulle vurdere samme oppgave. De uerfarne strøk alle basert på at studentene hadde feilet i forhold til kriteriene, hvor de erfarne vurderte mange av studentene som kompetente.

Målet er å unngå feil ved vurdering og inkonsistens, men både vurderernes bedømming og «compensation» blir brukt (Wolf 1995, s. 72). Wolf (1995, s. 133) presiserer at vurdererne ikke operer i et vakuum, men er en del av en sosial kontekst. Vurderere uttrykker at det er vanskelig å gjennomføre vurderinger som bare er basert på det objektive og standardiserte, og at det ikke bare kan ha en teknisk tilnærming.

Jones (1999, s. 154) viser til to faktorer som er vesentlige for vurderer, hvor den ene er knyttet til deres egen faglighet og yrke. Den andre handler om å ha kunnskaper om yrkesutdanning og vurderingsteknikker sett i forhold til bedømmelsen. Det handler om å gjøre vurdererne trygge på seg selv, at de får innsikt i rollen som vurderer. Jones uttrykker dette slik: «It is time to forget about trying to write «transparent» assessment criteria and to acknowledge that the role of a professional vocational educator is to make educational judgments» (Jones, 1999, s. 155).

Johnson (2008, s. 26) viser til at selv med trening av vurderere og velutviklede vurderingskriterier vil det være utfordrende å ivareta den konsistensen som bør ligge til grunn for de vurderinger som gjøres. Det er behov for å skape mer transparens i vurderingssystemene, og derigjennom skape mulighet for mer rettferdighet rundt vurderingene. Johnson presiserer at dette gjelder særlig ved bruk av helhetlig vurdering hvor vurdererne trekker inn flere faktorer fra forskjellige kontekster. Samtidig har deres egne verdier en betydning, og uttrykker «Another important consideration is the value system within which the reader/assessor exists and which affects their thinking» (Johnson 2008, s. 29). Samtidig viser Johnson (2008, s. 30) blant annet til Wolf (1995) som uttrykker at kriterier ikke kan dekke hele aspektet av ulike meninger. Det vil etterlates et rom for individuell tolkning hos vurdererne. Laming (2004, s. 164) viser til vurdererne som trekker inn egen erfaring hvor bedømmingen er usikker og uttaler at «... past experience enters like air rushing into a vacuum.» (Laming, 2004, s. 164). Basert på subjektiv erfaring kan vurderingene bli repetert i nye situasjoner og derigjennom gi økt selvtillit. Dette kan være en utfordring fordi

man kan havne i en situasjon hvor man ikke er fullt klar over nøyaktigheten i vurderingene og hvor feil kan gjøres. Dermed vil den repeterende vurderingen gi en systematisk feil.

O`Grady & Alwis (2012, s. 200) uttrykker at helhetlig vurdering avhenger av den profesjonelle bedømmingen til lærerne og «... implies a professional body to which the judgment can be ascribed» (O`Grady & Alwis, 2012, s. 200). Å samle inn nok informasjon ved summativ bedømming er vesentlig i henhold til de retningslinjer som foreligger. Dette informasjonsgrunnlaget utgjør de sentrale og avgjørende deler for hver enkelt student.

Vurdererne må bruke sin kunnskap om emnet og individet for å gjøre en rettferdig vurdering. I prosessen er det ikke bestemte prosedyrer som følges, men det legges vekt på å ta hensyn til hvordan de lærende på ulike måter uttrykker seg i sine framlegg. Betydningen av å være åpen er vesentlig for å se hva de lærende kan og viser fram. En helhetlig tilnærming til vurdering blir noen ganger kritisert fordi det ligger for mye subjektivitet i det, og kan være urettferdig fordi hver student kan bli vurdert forskjellig. Men det er nettopp dette som er styrken i helhetlig vurdering. O`Grady & Alwis (2012, s. 202) skriver:

Students who are assessed holistically are still judged against general criteria and the objectives of the problem but in addition, the measurement and fostering of understanding in respect to the uniqueness of each student and how he/she went about understanding is contextualized into the judgment. (O`Grady & Alwis, 2012, s. 202).

De som vurderer må se an hver enkelt situasjon for å finne ut hvordan de skal vurdere, og forstå det den lærende har prøvd å uttrykke gjennom sin forklaring.

Beckett og Hager (2002, s. 185) skriver at vurderinger som gjøres på arbeidsplassen er helhetlig, hvor vurderingspraksis er mer enn bare en teknikk. Profesjonelle bedømminger er helhetlige og involverer det kognitive, praktiske, moral, holdningene og følelser og er påvirket av det sosiale elementet. Med alle disse elementene vil det sjelden være vurderinger som sier noe om riktig eller galt, men de vil være med i den bedømmelsen som sier noe om riktigheten av det som skal vurderes. Helhetlige vurderinger på arbeidsplassen vil også være gjenstand for en kontinuerlig bedømming hvor mange mellomliggende vurderinger gjøres før man faller ned på den endelige vurderingen. Vurderinger gjøres også i forhold til kunnskap som er mer forankret til det institusjonelle ved utdanningen, og det som er knyttet til uformell læring i konteksten. Beckett og Hager (2002, s. 186) viser til vurderingskonteksten, der noe skal gjøres på rutiner, mens andre ting må gjøres ut i fra den spesielle konteksten handlingen skjer i. Videre må de som gjør vurderinger også tilpasse seg endringer som arbeidsplassen blir utsatt

for. Nye metoder og tilpasninger må til for å følge med på de fortløpende endrede mål og handlinger.

Dobson (2010, s. 32) viste til skjønnsbegrepet som en del av helhetlig vurdering. Molander og Grimen (2008, s. 179) har skissert det de kaller restkategori av skjønn, hvor skjønn kan ses som «... en uomgjengelig side ved en type praksis som anvender generell kunnskap, nedfelt i handlingsregler, på enkelttilfeller» (Grimen & Molander, 2008, s. 179). Man kan ikke ha allmenne handlingsregler hvor det forventes at det skal gi entydige konklusjoner. Dermed må slike beslutninger fattes ved å bruke dømmekraft. Hadde ikke det vært slik ville man kunne gjøre oppgaver mer eller mindre mekanisk, og hvem som utførte disse ville være uten betydning (Grimen & Molander, 2008, s. 179). Skjønnsutøvelse innebærer dermed en ubestemthet som må lukkes med skjønn. Det er ikke mulig å avgjøre ting uomtvistelig og skjønn må komme inn å fylle hullene. Dworkin (1978, s. 32) skiller mellom to typer skjønn, sterkt og svakt. Sistnevnte handler om at den som skal utøve skjønn er sterkt tilknyttet og bundet av standarder. I motsatt tilfelle er den som skal utøve sterkt skjønn ikke bundet av standarder. Ved skjønn innenfor standarder må dette omtales ovenfor den myndighet som har utformet disse. Utførelsen av skjønn kan dermed skje under to ulike omstendigheter. Den ene, hvor det ikke finnes lover og regler, baseres på tidligere eksempler. Den andre skal fylle tomrom og hull i de gjeldende lover og regler.

Det siste elementet ved forutsetninger for vurdererne er Wood (2014, s. 411), som viser til betydningen av førsteinntrykket og de underliggende kognitive prosessene og nøyaktigheten i disse. Når det gjelder kognitive prosesser knyttet til førsteinntrykket, viser Wood til system 1 (kognitive prosesser som er raske, ikke analytiske, automatisk og ubevisste). De som vurderer er ikke like klar over hvordan de fikk et førsteinntrykk. Dette skjer fort, og de er ikke bevisste den påvirkningen som skaper tilleggskrav. I forhold til førsteinntrykk og nøyaktighet, kommer det fram at det kan lede til feil i bedømmelsen og viktigheten av å være bevisst dette. Samtidig viser Wood til at prosesser som tar lengre tid ikke nødvendigvis gir riktigere bedømmelse enn de som skjer raskere. Å diskutere nøyaktighet er en utfordring. Det finnes ikke alltid en «gylden standard» (s. 415). Førsteintrykket kan påvirke vurderingen som en selvoppfyllende profeti. Denne kan være både positiv og negativ, og vil påvirke den videre vurderingen. Den kan også bidra til halo-effekten (s. 419), som gir vurderer et generelt inntrykk som blir med inn i vurderingen.



I gjennomgangen av teoretiske aspekter om helhetlig vurderinger knyttet til forutsetninger er det flere elementer. Det første er å gjennomføre holistiske vurderinger hvor både Wolf (1995), Johnson (2008) og Beckett og Hager (2002) trekker fram det som essensielt tilknyttet de handlinger som vurdererne utfører. En vesentlig faktor i dette arbeidet er det kontekstuelle i forhold til hvor disse vurderingene gjøres, noe som må tas hensyn til. Det neste er vurderernes handlinger og erfaring i vurderingsarbeidet. Her viser Wolf (1995) og Jones (1999) til faglighet og vurderingskompetanse, som blir en del av vurderernes forutsetninger. Laming (2004) viser til erfaringens betydning hos det enkelte menneske ved gjentatte vurderinger, noe som trekkes med inn i all vurdering og som kan utfordre kriteriene og nøyaktigheten. Samtidig peker Johnson (2008) på behovet for transparens, spesielt ved helhetlig vurdering. Det stilles spørsmål ved subjektiviteten i helhetlige vurderinger, men O'Grady og Alwis (2012) trekker fram nettopp dette som en styrke. Det unike ved hver vurdering blir i større grad ivaretatt. Skjønn i sterkere eller svakere grad er en del av det å gjøre helhetlig vurderinger, og Grimen og Molander (2008) viser til at dette blir en naturlig del ved det å gjøre bedømminger. Til slutt er det betydningen av førsteinntrykk i vurderingen som Wood (2014) peker på, og hvordan dette kan spille inn i vurderingene. I denne studien vil forutsetninger handle om det kognitive, verdiene og personligheten til den enkelte vurderer, samtidig som forutsetninger også knyttes til selve vurderingsprosessen.

Ut fra et epistemologiske syn om teorien rundt vurderernes forutsetninger kommer det frem et syn om både det konstruktivistiske og sosiokulturelle. Vurdererne operer i ulike kontekster hvor de må ta hensyn og tilpasse seg, noe som vil prege hvordan de vurderer. Skjønn vil blant annet spille inn når de gjør sine helhetlige vurderinger av en lærling og bli en del av et konstruktivistisk syn på vurdering.

### 3.1.9 Oppsummering teoretisk rammeverk for helhetlig vurdering

Under hvert avsnitt (3.1.3-3.1.8) gjorde jeg en oppsummering av de teoretiske aspektene om helhetlig vurdering og som utgjør grunnlaget for diskusjoner i kapittel 5 og til dels i kapittel 6.

Wolf (1995) sin redegjørelse om kompetansebasert vurdering vil være sentral. Hennes innledende definisjonen om kompetanse basert vurdering framstår for meg som en idealistisk tanke om en ønsket situasjon, og blir gjennom Wolf sin videre diskusjon stilt i et kritisk lys.

I den sammenheng har jeg utviklet min egen definisjon på helhetlig vurdering:

Helhetlig vurdering er å trekke inn ulike aspekter av en handling knyttet til en yrkesutøvelse i en konkret kontekst, der vurderer bruker sin erfaring, kompetanse og skjønn for i størst mulig grad å vurdere yrkeskompetansen opp mot de krav som stilles for utøvelsen av et yrke.

Definisjonen må ses på som et forsøk på å beskrive en kompleks situasjon hvor man skal ivareta ulike sider ved en kompetanse samtidig som hensynet til den aktuelle konteksten er vesentlig å ivareta. Definisjonen kan sies å rette søkelyset mot innholdssiden i vurderingen og i mindre grad den metodiske siden. Jeg vil komme tilbake til det senere i studien.

I kapittel 3.1.1 antydet jeg at alle de fire epistemologiske retningene ville kunne være representert i vurderingen av en yrkeskompetanse, uten at det ble redegjort for i hvilken grad. I gjennomgangen av valgt teori om helhetlig vurdering ble hvert aspekt belyst om hvordan de fire epistemologiske retningenes representasjon framkom. Overordnet er det nærliggende å si at i det teoretiske rammeverket vil både den konstruktivistiske og sosiokulturelle være de mest fremtredende, og i mindre grad den instrumentelle og disiplinorienterte. De to sistnevnte kan hovedsakelig kjennetegnes ved faglig innhold og mål og kriterier.

Gjennomgangen i kapittel 3 danner det teoretiske grunnlaget for hva som ligger i helhetlig vurdering, og som skal bidra til å finne ut i hvilken grad vurdererne gjør helhetlig vurdering og hva som kan påvirke de. Avslutningsvis presisere at valgt teorigrunnlag ikke ses på som ferdig utviklet for tematikken helhetlig vurdering. Valgene vil også være farget av mitt eget epistemologiske syn om kunnskap og vurdering.

### 3.2 Forskning på vurdering i en internasjonal og norsk kontekst

Forskning på feltet om vurdering er ikke omfattende innen fag- og yrkesopplæring, spesielt ikke innen helhetlig vurdering. De studiene som presenteres fra Norge og internasjonalt har ikke direkte fokus på helhetlig vurdering, men er studier om vurdering hovedsakelig i det yrkesfaglige feltet hvor det kommer funn som vil være interessant i forhold til helhetlig vurdering. De utvalgte studiene vil bli presentert hovedsakelig i følgende struktur: studiens formål og de spørsmål som stilles, metodiske løsninger og resultater og sentrale funn.

#### 3.2.1 Forskning på vurdering i en internasjonal kontekst

I et internasjonalt perspektiv er ikke forskningen om vurderernes vurderinger omfattende relatert til fag- og yrkesopplæringen. Siden et av vurderingssystemene benyttet i HELVYRD-

studien er et «large-scale testverktøy» vil jeg først redegjøre for noe av bakgrunnen og utviklingen i denne delen av vurderingsfeltet. Deretter vil tilhørende forskning på KOMET/MECVET-modellen bli redegjort for. Deretter følger en kort redegjørelse for to avhandlinger gjort i Sverige og Danmark. Til slutt kommer en redegjørelse for to studier, en gjennomført i et yrkesfagmiljø, den andre i en allmenn skole.

Det europeiske kvalifikasjonsrammeverket (EQF) har som intensjon å systematisere rammer for kompetanse i form av læringsutbytter. Dette gir muligheter for å sammenligne på tvers av institusjoner i samme land og mellom land. Dette har resultert i nasjonale kvalifikasjonsrammeverk som er tilpasset de respektive lands behov. Det er gjort forsøk på å utvikle instrumenter for sammenlignende målinger av yrkeskompetanse innenfor VET. Et tysk prosjekt kalt PISA for VET ledet av Martin Baethge med medforfattere ga ut en bok i 2006, *PISA-VET: A Feasibility-Study*, hvor resultatene av prosjektet ble presentert. Senere ble det tatt et initiativ for VET-LSA som ble gjennomført i en internasjonal sammenheng (Baethge & Arends, 2009, s. 8). Norge deltok i dette forsøket hvor det ble levert en rapport med blant annet elektrikerfaget Olsen (2009, s. 6). Flere fag var valgt ut blant annet elektrikerfaget. Intensjonen med VET-LSA var å utvikle en storskala-test i samarbeid med en rådgiverkomite for ovennevnte PISA-VET. Det deltok forskere fra åtte land. Her ble grunnlaget og «key-aspects» for å utvikle VET-LSA. O\*NET databaser ble brukt for å definere yrkesprofiler som et grunnlag for å lage testoppgaver. I sluttrapporten nevnes hva som måtte gjøres for å videreutvikle VET-LSA. En av anbefalingene var å inkludere elementer av praksis i testene, siden dette er av stor betydning innen VET. Det var også behov for å styrke og å lage et rammeverk som ytterligere beskrev de yrkesoppgaver som gjaldt, kvalifikasjoner og kjerneområdene i yrkene.

Resultatene fra rapporten til Olsen (2009, s. 7) viser blant annet til vurdererne i elektrikerfaget. Det slås fast at fullverdige vurderinger av fag- eller yrkeskompetanse bare kan gjøres i situasjoner hvor konkrete praktiske arbeidssituasjoner oppstår.

Når prøvenemndsmedlemmer gjør sine vurderinger, foretar også de sammenligninger, som alle andre eksaminatorer, men de bygger altså på en lang rekke kriterier, mer eller mindre eksplisitte, basert på formulert eller taus kunnskap, bygd opp gjennom erfaringer og kommunikasjon i faglige institusjoner og fellesskap. (Olsen 2009, s. 7)

Det vises til kompleksiteten i det å gjøre vurderinger, samtidig må man se på nytten og funksjonen til ulike vurderingssystemer og hvilken hensikt disse har i en

vurderingssammenheng. Det presiseres at slike tester kunne ha en funksjon knyttet til enkelte sider av det å vurdere en fag- eller yrkeskompetanse, og dermed bidra til en sammenligning. VET-LSA-prosjektet ble ikke videreført, men har gitt erfaringer som har lagt grunnlag for å diskutere og utvikle andre forsøk.

Med Rauners et al. (2013, s. 157-163) utvikling og utprøving av KOMET-modellen (teoretiske grunnlaget belyst i kapittel 2.2 og kapittel 3.1.6) har gitt erfaringer knyttet til vurderinger i forbindelse med treningen av ratere i Kina.

Rauner et al. (2013, s. 157) viser til erfaringer fra raternes vurderinger. Innledningsvis hadde raterne en tendens til å gi for gode score. Grunnen var at en analyse av oppgaveløsningen ble startet uten å anvende løsningsrom. Dette medførte, ifølge KOMET-prosjektet, at raternes egne forventninger og holdninger ut fra egen undervisning og treningspraksis fikk en dominerende rolle i ratingprosessen. I denne fasen varierte også resultatene mye. Dette hadde sammenheng med at forståelsen av systemet for vurdering ennå ikke var velutviklet. I de videre fasene ble det lagt vekt på at raterne skulle få mer innsikt i bruk av items (kriteriespesifikasjoner) og løsningsrom, og at dette i større grad skulle sikre en mer standardisert rating prosedyre. I de påfølgende rundene med rating av besvarelser oppdaget KOMET-prosjektet at raterne ble mer kritiske og at dette var knyttet til det begrensede antall items, dessuten at løsningsrommet var en guide som ga innspill til ulike løsninger, ikke bare til en ideell løsning. Etter fire gjennomførte rater-seanser økte «interrater reliabiliteten» betraktelig. En vesentlig faktor var styrken i det å dele ratingen med en eller flere ratere, og den omforente ratingen hadde stor effekt på resultatet.

Rauner (2014, s. 7) har gjennomført en sammenligning mellom den offisielle tyske eksamen innen yrkesfag og KOMET-modellen. Han viser til at de kompetanseorienterte og helhetsorienterte vurderingsmetoder knyttet til eksamen konseptuelt ikke skiller seg mye fra prosedyrene rundt ratingen i KOMET. Skillet er mest i gjennomføringen, hvor KOMET har den mer konkrete og nøyaktige måten å gjøre rating på. Dette har sammenheng med deres tydelige kriterier og prosedyrer, i motsetning til eksamen som er mer knyttet til en mangfoldig og praktisk erfaring og tilnærming. I Rauner (2014, s. 17) sine analyser og sammenligninger av metodene slås det fast at de dimensjoner og kompetansenivåer som er utviklet i modellen kan forankres i de opplæringsrammene som gjelder for de offisielle eksamenene i Tyskland. Det er i stor grad også samsvar mellom formatet på de holistiske eksamensoppgavene og testoppgavene i KOMET (s. 20). Formatet på de helhetlige oppgavene er sett på som en

styrke. Gjennom bruk av to ratere pr. oppgave vil interrater-reliabiliteten øke. Samtidig påstås det at KOMET-modellen har høyere grad av gyldighet på testresultatene sett i forhold til de mer konvensjonelle testene, og de kompetansenivåer og kompetanseprofiler som blir oppgitt (s. 31). I avhandlingen til Jacobs (2016, s. 166) kom det fram at et stort antall (92%) mente oppgavene brukt i KOMET-modellen skulle bli en del av «curriculum», og innlemmes der. Samtidig kunne nesten alle (94,1%) relatere disse til sitt eget yrke.

To nordiske avhandlinger er gjennomført på vurderingsprosessen innen fag- og yrkesopplæring. Tsagalidis H. (2008, s. 19) har i sin avhandling sett nærmere på bedømminger som gjøres av emnelærere innenfor Hotell- og Restaurantprogrammet i den svenske yrkesopplæringen. Hovedfokus er å se på hva vurdererne vurderer av elevenes «kunnande» innenfor dette yrkesprogrammet. Forskningsspørsmålene er rettet inn mot kunnskapsformer, innhold og kjennetegn. Metodisk (s. 73) er det en kvalitativ studie med innsamling av data i 6-7 år. Aktivitetene i kjøkken har blitt filmet, dessuten har det blitt gjennomført intervjuer og refleksjoner i grupper og enkeltvis. Noen av de overordnede funnene Tsagalidis (2008, s. 189) kommer fram til er at vurdererne legger vekt på både praktisk arbeid og teoretisk kunnskap, samt passende handlinger i arbeidet. Videre kom det fram i intervjusituasjonene at vurdererne reflekterte mer på de overordnende målene og kriteriene. Dette står i kontrast til det som fant sted i den konkrete vurderingssituasjonen, hvor vurdererne fokuserte mer på den konkrete måten de studerende løste oppgavene på. En konklusjon er at vurdererne kjente til det overordnende vurderingssystemet, og at dette ble tilpasset vurderingssituasjonen. Vurderingen baseres også på deres egen kunnskap i den yrkesmessige kulturen. Det kan oppfattes som en ulempe for studerende at de i større grad blir utsatt for kriterier knyttet til hvordan det reelle yrket utføres enn de mål og kriterier som er lagd i skolesystemet. Et siste funn viser mangler hos vurdererne innen det enkelte fag. De kan ikke se bort fra de yrkesmessige kunnskaper og ferdigheter og den kulturen som eksisterer i yrket når de gjør sine vurderinger.

I Friche (2010, s. 11-13) sin avhandling er det sett nærmere på evalueringspraksisen i de danske erhversskolene (yrkesskolene) med fokus på intensjoner, anvendelse og virkning. Forskningsspørsmålet er rettet inn imot hvordan evaluering i nasjonale utdanningspolitiske og erhversskolelære får gjennomslag hos faglærere og elever som erhvervsfaglige praksisformer. To underspørsmål er knyttet til dette. Det ene setter lys på hvilke diskurser som er gjeldende for evaluering i erhvervsutdannelse. Det andre handler om hvordan faglæreres konstruksjon av utdanning og undervisning kan forstås og forklares. Metodisk har

Friche (2010, s. 78) valgt multiple strategier i forskningsdesignet hvor det primære er å samle inn data ved hjelp av kvalitative forskningsintervjuer med observasjoner og gruppeintervjuer som utfyllende data. Dokumentstudier og sekundærdata fra andre undersøkelser har også fylt ut bildet. Informanter har vært elevgrupper, lærergrupper, den enkelte lærer og den enkelte leder med fokus på følgende fag: smedutdanning, gastronomutdanning, tannklinikkassistentutdanning og mediegrafikerutdanning. Når det gjelder Friche (2010, s. 355) funn, er disse systematisert i tabeller og tekst. Det er for omfattende å gå i detalj på disse her, men flere av funnene er gjenkjennbare både når det gjelder formative og summative aspekter rundt vurdering. Det overordnede er at Friche viser til eksempler på ulike vurderingspraksiser i forskjellige yrker, og hvordan disse er påvirket av ulike faktorer i ulike kontekster og sosiale situasjoner. I en senere artikkel viser Friche (2013, s. 17) til at ved et studium (smedutdanning) ble evalueringspraksis «... formet af omverdenes kulturelle konstruktion af en produktorientert logik ...» (Friche 2013, s. 17). Her uteble den prosessorienterte vurderingen, dermed ble det ikke en helhetlig vurdering. Dette er en del av diskusjon om bærekraftig formativ og summativ vurdering sett i forhold til en livslang læring, noe som også kan knyttes til Tanggaard og Elmholdt (2008, s. 113) sin påpekning om viktigheten av sammenheng i vurderingsfasene.

Johnson (2008, s. 26) har fokusert på vurderernes kognitive strategier på helhetlige bedømmelser av portefolier innen helsearbeiderfaget i England. Samtidig ble de sosio-kulturelle faktorene studert. To hovedspørsmål ble knyttet til denne studien. Det første handlet om hvilke elementer av helhetlige beskrivelser vurdererne støtter seg på i sin vurdering av portefolier. Det andre spørsmålet var knyttet til hvilke faktorer som kan påvirke vurderernes bedømmelser i ulike situasjoner (s. 32).

Metodisk (s. 32) ble forskningen gjennomført i flere faser med seks vurderere. Den første fasen var knyttet til den kognitive delen og handlet om å samle inn data ved bruk av «think-aloud» i vurderingen av portefolien med støtte av en intervjueteknikk (basert på en modifisert Kelly's Repertory Grid (KRG)). Den andre fasen handlet om innsamling av kvalitative data knyttet til den sosiokulturelle delen hvor tre forskjellige skoler ble observert i gjennomføringen av vurderingsarbeidet. Dette la i sin tur grunnlaget for å utvikle et semi-strukturert intervju som ble gjennomført på hver av vurdererne etter vurderingen av portefolien.

Resultatene av undersøkelsen (s. 33-37) er delt inn i en kognitiv del med think aloud og KRG data, samt en del med fokus på hva som påvirker konsistensen i bedømminger. Et funn var at

vurdererne brukte kriteriene i deres gjennomgang av portefoliene, og at de var påvirket av strukturen på hvordan disse var bygd opp. Videre kom det fram at vurdererne balanserte den helhetlige bedømmingen ved å legge mindre vekt på enkelte kriterier, mens andre kriterier fikk større betydning. Videre viste det seg at det var vanlig blant vurdererne å bruke annen informasjon utenfor kriteriene for å bedømme kvaliteten. I denne sammenheng var det også vurderere som brukte referansepunkter utenfor portefolien som støttet i bedømmelsen. De kommenterte også rollen til de som hadde veiledet elevene, og hvordan dette kunne ha påvirket arbeidet, hvordan dette må tas hensyn til i vurderingsarbeidet og komme elevene til gode. Til slutt i denne runden ble det pekt på hvilken betydning det har i vurderingsarbeidet å utvikle felles forståelse av nøkkelbegreper i faget. Videre trekker Johnson (2008, s 37- 38) fram fire områder som er med på å påvirke vurderingene. Den første handler om vurderernes syn på den lærende i vurderingsprosessen. Man må ha et system tilpasset den lærende hvor man ser det helhetlige, et system som kan få fram den enkelte elevs framgang i læringen. Det andre elementet er betydningen av å bygge gode relasjoner til andre vurderere. Det tredje aspektet er at det systemet som blir benyttet kan fremme den forventede profesjonalitet, og at dette er noe man kan stole på. Vurdererne må kunne gi sin mer personlige profesjonelle vurdering. Det siste området handler om det å ha forpliktelse til sitt eget fag, og at vurdererne har kontakter og forbindelser til yrket for å holde seg oppdatert.

Crisp (2012, s. 10) har gjort forskning på vurderernes kognisjon i en kontekst hvor man vurderer skole-baserte prosjekter for «high-stakes» formål. Studien er ikke innen fag- og yrkesopplæring men har interessante og relevante aspekter ved seg. Studien viser til mye forskning som er gjort på feltet, men argumenterer med at det er mangler forskning på vurderernes vurderinger i forhold til større arbeider som prosjekter osv. Noe av argumentasjonene til Crisp ligger i følgende sitat:

It is sometimes assumed that the human scoring process is a simple matter of applying the assessment criteria to a response, but the lack of perfect agreement between raters in many educational assessment contexts would indicate otherwise. Without a clear understanding of the process involved in scoring and of how the scoring criteria must be resolved against an assessor's intuitive professional impression of a piece of work, can we ever really be sure of what an assessment is measuring? (Crisp, 2012, s. 10).

Metodisk (s. 12-14) var dette en studie gjennomført med “think-aloud” (verbal protocol) på tretten lærere og tre moderatorer som gjorde sine vurderinger knyttet til fagene IKT, geografi

og engelsk litteratur. Crisp har også plukket ut tre moderatorer som deltok i undersøkelsen. Alt ble tatt opp på bånd og transkribert. I analysen ble det utviklet koder for den transkriberte teksten, som til dels var influert av kodelister utviklet i andre tidligere studier. I denne prosessen valgte de også å involvere deltakerne til å se på kodene og komme med innspill og se dette som en del av valideringsprosessen. Jeg vil presentere kodingskategoriene de har endt opp med i denne studien, og vise til noen av de resultatene som kommer fram i denne studien (s. 17).

*Den første kategorien* handler om planlegging og orientering, hvor vurdererne bruker tid på å sette seg inn i studentens arbeide og hvordan de skal gå fram i vurderingsarbeidet. Her kan de også lage kommentarer til hva de venter å se i besvarelsen. *Den andre kategorien* handler om det å lese og forstå arbeidet og hvordan vurderer gjør dette. Dette kunne gjøres ved å skanne gjennom arbeidet for å få en oversikt over strukturen og viktige deler av innholdet før man leste dette mer utførlig. Vurdererne kunne også sitere og oppsummere fra studenters arbeid og utvikle deres egen mentale forståelse av dette. *Den tredje kategorien* handler om realisering av oppgaven og hvordan vurdererne opplever at studenten har gjennomført dette. Det var to koder knyttet til dette som handlet om kommentarer på lengden av arbeidet og på fokuset mot det oppgaven skulle handle om. *Fjerde kategori* handler om sosiale og emosjonelle reaksjoner fra vurdererne når de gjør deres vurderingsarbeid. Ulike reaksjoner ble registret i forhold til studentens arbeid. Disse kunne handle om alt fra glede til å mislike det de leste. Det forekom også refleksjoner på egen vurdering av studenten. Vurdererne kunne også « snakke med » studenten eller om studenten. *Den femte kategorien* handler om å vurdere samtidig eller underveis. Dette vil si at vurdererne gjorde vurderinger fortløpende mens de leste hvor det under lesing ble gjort sammenligninger av deler av arbeidet. Det ble gjort notater på positive og negativ sider ved studentens arbeid fortløpende, og det ble gjort kommentarer på ting som manglet. *Den siste kategorien* handlet om en overordnet vurdering til slutt. Denne innebar å gå gjennom vurderingen av arbeidet, se enkeltdelene, komme til en avgjørelse og ta i betraktning den endelige vurderingen. Crisp skriver:

This process usually involved using the scoring criteria and summarizing the evaluations made so far, including strengths and weakness, often in relation to specific criteria and perhaps also through use of their internalized understanding of the scoring criteria, and/or their mental representations of examples of work they have previously seen. Assessors often flicked back to parts of a project when summarizing and justifying their judgments. (Crisp, 2012, s. 17).



Det påpekes at vurdererne ofte kan gjøre refleksjoner i forhold til den overordnede vurderingen, og om dette er i tråd med de forventninger og følelser man hadde til kvaliteten på det arbeidet som ble bedømt.

Det sentrale fra internasjonal forskning vil bli oppsummert i kapittel 3.2.3. Men først en gjennomgang av den norske forskningen.

### 3.2.2 Forskning på vurdering i en norsk kontekst

Når det gjelder forskning på vurdering i Norge innen fag- og yrkesopplæring er det fire studier jeg vil trekke fram. Disse er knyttet til prøvenemndene og fagprøven. Forskning på den skriftlige eksamensformen finnes ikke i Norge, men det kan knyttes til Crisp (2012) sin forskning.

I forskning på fagprøven vil jeg starte med Engelsen (1998, s. 183-199) som i sin studie har brukt litteraturstudier og dokumentanalyse for å belyse ulike sider av fagprøven som en evalueringsform innen fag- og yrkesopplæring. Engelsen har forsøkt å beskrive evalueringsbegrepets innhold etter noen sentrale dimensjoner formulert som spørsmål. Disse kan være relevante for å belyse viktige elementer ved alle vurderingssystemer. I den første dimensjonen stilles spørsmålet om hva formålet med evalueringen er. Det neste spørsmålet handler om hva som skal være gjenstand for evaluering. Videre stilles spørsmålene om hvilke kriterier som skal ligge til grunn for evaluering, til hvilken tid og hvem evalueringen er til for. De to neste spørsmålene er knyttet til hvem som skal utføre evalueringen og hvordan dette skal skje. Det siste spørsmålet handler om hvilken skala dette skal skje etter. Det blir sagt at det trengs at man klarlegger vurderingskriteriene for kandidatene, noe som er et sentralt element, selv om dette er i utvikling de senere årene gjennom reformer og lovgivning. Det blir også sagt at det er behov for en valid og reliabel prøvesituasjon, åpenhet rundt fagprøven og det å bli en del av det offentlige rommet. Engelsens spørsmål og påpekninger er viktige bidrag i den videre diskusjonen om vurdering.

FAFO-rapporten fra Deichman-Sørensen et al. (2011) hadde tre sentrale problemstillinger (s. 12-13) knyttet til forskningen om prøvenemnder og gjennomføring av fag- og svenneprøven. Det første var gjennomføringen av fag- og svenneprøven med underspørsmål om vurdering, læreplanmål, prøveforskrifter, oppfatning av eget fag, likheter og forskjeller. Det neste spørsmålet er knyttet til handlingsrommet, hvordan dette forstås og utnyttes i forbindelse med fag- og svenneprøver. Det tredje spørsmålet handler om identiteter sett i forhold til rollen i en

prøvenemnd, hva det vil si å være en offentlig prøvenemndsmedlem sett i forhold til det å være en faglig ansatt i arbeidslivet. Dette handler også om deres faglige kompetanse, om å ta vare på denne innen det gitte handlingsrommet.

Forskerne har i sine undersøkelser (s. 14-15) gjennomført fokusgruppeintervjuer med prøvenemndsledere eller andre medlemmer i fem forskjellige fag inkludert elektrikerfaget. Det kan være av betydning hva som også kommer fram i de andre fagene. I elektrikerfaget deltok tre personer fra tre fylker for å sikre en geografisk spredning og representasjon fra ulike næringsstrukturer. Fokusgruppeintervjuene varte fra tre til fire timer, hvor to til tre forskere deltok. Disse ble tatt opp på bånd og transkribert i fulltekst. I intervjuet ble det lagt vekt på å få fram det som faktisk skjer i gjennomføringen av fag- og svenneprøver. Det ble ikke stilt spørsmål som kan låses til de gjeldene retningslinjer og bestemmelser. Det må understrekes at datagrunnlaget i denne undersøkelsen er begrenset og ingen generelle konklusjoner kan trekkes.

Sentrale funn fra vurdering knyttet til elektrikerfaget (s. 60-62) viste at ved bestått fagprøve måtte produktet være salgbart for kunden og at el-sikkerhet ble ivaretatt. Planlegging og dokumentasjon samt bruk og forståelse av forskrifter ble sett på som viktig. Ved fordeling av karakterer var det variasjoner mellom fylkene når det gjelder stryk og de to andre karakterene. Samtidig var det lite uoverensstemmelse i nemndene ved stryk eller meget godt bestått. I karakterskalaen var det et ønske om ytterligere et nivå i karakterskalaen siden karakteren bestått opplevdes som for bred, at forskjellen mellom de som får denne karakteren ble for stor. Når det gjelder skillet mellom den kritiske karakteren ikke bestått og bestått, går dette hovedsakelig ut på mangler knyttet til el-sikkerhet. Når det gjelder vurderingskriterier, oppleves det å være ulik grad av presisering og praksis for å informere kandidatene om grunnlaget for vurderingen.

Når det gjelder faglig bedømmelse, er det ingen tvil om synet på den rollen vurdererne har som prøvenemndsmedlemmer, selv om den faglige identiteten kan variere imellom fagene. Det er fagpersonen og fagpersonens eget skjønn som ligger til grunn for de bedømmelser som gjøres. «Dette skjønnnet bygger på egen erfaring, etablerte normer og diskusjon og refleksjon om fagets utvikling.» (Deichman-Sørensen et al 2011, s. 76). Undersøkelsen viser at kandidatens utprøving skal ta utgangspunkt i faglig forståelse og helhetlig kompetanse. I omtalen av en god kandidat og en god prøve vektlegges det helhetlige. Rapporten uttrykker at «Prøvenemndene utøver et skjønn basert på en helhetlig vurdering av det konkrete arbeidet

som gjøres, ikke basert på et sett med generelle kriterier løsrevet fra komplekse og særegne oppgaver» (Deichman-Sørensen et al., 2011, s. 77). Det er ingen tvil om at det er prøvenemndsmedlemmene selv som mener at de skal bedømme den faglige kompetansen. Forslag til videre forskning handler om å se på prøvenemndenes vurderingsfaglige kompetanse, om denne er tilfredsstillende. Det kan være ulike vurderingskulturer i skolene eller i de instansene som gjennomfører vurdering av faglig standard.

Nyen og Tønder (2014, s. 121) har i sin bok kommentert FAFO-rapporten. De peker på at prøvenemndene har en viktig rolle i vurdering av fagkompetanse, men at denne rollen ofte får lite fokus. Fagfolkene som utgjør prøvenemndene gjennomfører sine vurderinger på premisser som er arbeidslivsrelevante. Det finnes et felles sett av normer for å vurdere hva som kjennetegner et godt fagarbeid. Prøvenemnda bestemmer om kandidatene har den ferdighet og kunnskap som kreves. Læreplanen er et grunnlag for bedømmingen, men de faglige normene er også uavhengige og står ikke uttrykt nødvendigvis i læreplanenes terminologi.

Nore og Lahn (2015, s. 145) har i sin artikkel i sluttrapporten til Kvalitet i fag- og yrkesopplæring tatt for seg spørsmål knyttet til hvilke kriterier prøvenemnder bruker for å vurdere kandidatens faglige nivå. Videre stiller de spørsmålet om hvordan faglig skjønn og et standardoppsett (vurderingsskjema, kriterielister m.m) spiller inn i vurderingen. De spør også om nemndene tar høyde for generiske ferdigheter og hybride kompetanser. Undersøkelsen har ikke tatt med elektrikerfaget men fire andre fag. Det er interessant å se på resultatene i forhold til noen generelle trekk.

Metodisk (s. 146) har det vært gjort dokumentanalyse av oppgavetekster, vurderingsskjemaer og kandidatenes dokumentasjon fra prøvene. Det ble utviklet kategorier for analyse av dokumentene som gikk på type oppgaver, vedlegg til oppgavene, informasjon til kandidater, krav til dokumentasjon og hva som faktisk ble dokumentert. Dokumentanalysen var et utgangspunkt for å gjennomføre semistrukturerte telefonintervju med totalt 15 ledere fra de ulike nemndene. Lengden på intervjuene oppgis ikke. Intervjuene ble gjennomført etter prinsipp om stimulert erindring (Engelsk: stimulated recall), hvor både intervjuer og informant satt med de aktuelle dokumentene foran seg. Det opplyses ikke når prøvene ble gjennomført og vurdert.

I forhold til oppgaver fant de ut at det i alle nemndene ble lagt stor vekt på å utarbeide oppgaver som var realistiske. Likevel ble det oppdaget stor variasjon i oppgavetyper både

innen lærefagene og mellom dem. Dette gjaldt fra standard til full skreddersøm, fra åpne komplekse til løsrevne deloppgaver (s. 149).

Når det gjelder utforming og bruk av vurderingskriterier viser studien til merknader i forskrift, hvor Utdanningsdirektoratet peker på at det ikke er tilstrekkelig å utøve faglig skjønn og å vurdere ut fra egne erfaringer fra arbeid (s. 157). For å sikre en objektiv og rettferdig vurdering må de ulike sidene av prøvearbeidet vurderes ut fra et utarbeidet grunnlag. I gjennomgangen av vurderingskriterier brukte de ulike nemndene i liten grad læreplanmålene som utgangspunkt. De tok mer utgangspunkt i nøkkelkvalifikasjoner.

I diskusjonen om utvikling av kriterier og skjema (s. 157-160) trekkes andre faktorer inn som kan innvirke på vurderingen. Disse viser blant annet til det helhetlige blikket knyttet til «førsteintrykket». I prøvenemndsarbeidet oppstår det ved å hilse på lærlingen første dag, noe som kan påvirke andre deler av vurderingen. Forholdet mellom faglig skjønn og det å vurdere med henblikk på vurderingskriterier relateres til en faglig praksis som er utviklet i vurderingsfelleskapet, hvor man ser etter bestemte ting, og hvor man lærer hva som skal vektlegges.

Når det gjelder bruken av karakterskalaen (s. 160) og begrunnelsene som ligger bak karakterene, finner forskerne at karaktersettingen ikke følger nøye regelverk, læreplaner og kompetansemål. «Argumentasjonen deres er mer knyttet til om det er godt nok til at de ville ansatt personen som faglært i eget firma.» (Lahn og Nore, 2015, s. 161)

De oppsummerer noen overordnede punkter om hva prøvenemndene i de ulike fagene faktisk la vekt på i vurderingen, og fant stor variasjon (s. 162). Når det gjelder vurderingsarbeidet stilles det spørsmål fra forskerne om hvor mye tid nemndsmedlemmene bruker på å sette seg inn i lærlingenes arbeid. Spørsmålene er knyttet til hvordan de kan vurdere kundebehandling, HMS, selvstendighet og problemløsningsevne og kommunikasjon og samarbeid med kunden og kollegaer.

Studien påpeker at alle vurdererne la vekt på å gjøre en helhetlig vurdering av prøven og kandidaten. For flere prøvenemnder innebærer det å vurdere mer enn bare ren fagkompetanse. Nøkkelkvalifikasjoner og generell del av læreplan blir også trukket inn. De er ikke nivået på en eksplisitt fagkompetanse de vurderer, men uttrykker at det «... ser ut til å være en blanding av skjønn basert på lang erfaring i prøvenemnd, ulike forventninger til kvalitet på utført arbeid

og ulike tolkninger av hva som kreves i henhold til læreplanene». (Lahn og Nore, 2015, s. 164)

Det konkluderes med at nemnder er en egen institusjon med egen vurderingspraksis. En sammenheng i vurderingspraksis kan sikres ved at representanter fra flere deler av fag- og yrkesopplæringen blir medlemmer i nemndene.

Rafoss og Witsø (2014) har i sin artikkel presentert en kvantitativ undersøkelse gjennomført på prøvenemnder i Agder. De har to hovedspørsmål (s. 4) hvor det ene handler om i hvilken grad prøvenemndene følger Opplæringslova i gjennomføringen av fag- og svenneprøver. Det andre spørsmålet er hvilke faktorer som påvirker prøvenemndene å følge Opplæringslova. Forfatterne viser til forskning på feltet hvor Friche (2010) og Deichman-Sørensen et al. (2011) ses på som to unntak i forskningen på prøvenemnder.

Metodisk (s. 9-10) har de gjennomført en kvantitativ studie av 225 prøvenemndsmedlemmer fra Vest-Agder samt noen fra Aust-Agder. Det ble delt ut ett spørreskjema til medlemmene i forbindelse med et opplæringskurs i vurderingsbestemmelsene i lovverk og læreplaner.

I sin gjennomgang av forskning og teori har forskerne, basert på deres forforståelse, utarbeidet fire hypoteser. De belyser potensielle spenninger mellom et formelt lovverk som Opplæringslova, som regulerer vurderingene til prøvenemnda, og vurdering preget av både faglig skjønn og kyndighet, eksplisitt og taus kunnskap. De ser to sider av dette, den ene er representant for fagene som ønsker autonomi og frihet, den andre er rekken av formelle krav fra fylkeskommunene (s. 8). De uttrykker at dette kan bli et dilemma for nemndene når disse skal vurdere et arbeid på grunnlag av generelle mål og vurderingskriterier, i vurderinger basert på faglig skjønn og egen praksis, og oppgaver som kan både være konkrete og abstrakte.

Jeg vil presentere kort de fire hypotesene forskerne har utviklet med tilhørende funn. Den første (s. 11) tar for seg spørsmålet om det er handlingsdelen av lovparagrafene som blir fulgt. Dette får en viss bekreftelse, og viser at det er behov for klare kriterier for karaktersettingene. Det andre (s. 12) handler om at økt alder og erfaring hos sensorene faller sammen med at lovens paragrafer ikke følges til punkt og prikke. Lang erfaring er ikke ensbetydende med at man følger opp lovparagrafene nøyere. Blant annet begrunnes dette med at man kan ha en faglig sikkerhet og selvstendighet som brukes i vurderingen av lærlingen. Den tredje (s. 12) er at kurs og prøveordninger ikke vil ha den effekten som er tenkt. Medlemmene ble spurt om de

hadde forslag til forbedringer, hvor de uttrykte ønske om å få større innflytelse på utformingen av vurderingskriterier og at de i større grad må tilpasses de ulike fagenes egenart. Den siste handler om de eldre og nyere medlemmene i fagene, om de førstnevnte følger loven i mindre grad (s. 13). Det kommer fram at det er stor spredning i svarene når det gjelder hvorvidt medlemmene følger lovpåleggene. Forskjellene fagene imellom er her ikke så store som andre studier viser til.

Sentrale funn fra gjennomgangen av de norske studiene oppsummeres i kapittel 3.2.3.

### 3.2.3 Oppsummering om tidligere forskning

I gjennomgangen av tidligere forskning på feltet ble det fokusert på studiens spørsmål, metodiske opplegg og til slutt sentrale resultater og funn. Her følger en kort oppsummering av de mest sentrale elementene med relevans for HELVYRD-studien.

De ulike studienes spørsmål påvirket forskningsspørsmålene i HELVYRD. Gjennomgangen viser forskning som har fokusert på vurderernes bedømminger innen ulike vurderingsformer, både formative og summative. Dette gjelder for eksempel fagprøve, tester, prosjekt og portefolier innen ulike fagområder. Det har bidratt til at HELVYRD-studien handler om flere vurderingssystemer. Samtidig har den blitt spisset inn mot et fag og et nivå i utdanningen. Hensikten har vært å komme mer i dybden om vurderernes helhetlige vurderinger.

I studienes metodiske framgangsmåte er det hovedsakelig kvalitative studier, men også et par kvantitative. Kvalitative intervjuer med vurdererne der «stimulated recall» og «think-aloud» blir benyttet, vil være metoder jeg tar med inn i designet av HELVYRD-studien. I tillegg vil dokumentanalyse være et bidrag i datainnsamlingen.

I studienes sentrale funn om vurdererne vil jeg strukturere det etter de samme aspektene redegjort for og anvendt om helhetlig vurdering i kapittel 3.1. Disse vil bli benyttet i diskusjonen i kapittel 5 og 6.

Om funn knyttet til *oppfatninger av vurdering* uttales det blant annet at vurderingen skal ta utgangspunkt i kandidatens faglig forståelse og helhetlig kompetanse, det helhetlig skal vektlegges. Engelsen (1998) understreker bevissthet rundt formålet med en evaluering, og viser til betydningen av en åpenhet rundt fagprøven i det offentlige rom. Deichman-Sørensen et al. (2011) hevder at i en vurdering er det den faglige forståelsen og helhetlige kompetansen som legges til grunn, og at vurdererne bruker skjønn. Lahn og Nore (2015) uttaler at nemnder

er en egen institusjon med sin egen vurderingspraksis. Skjønnsbasert vurdering knyttes til ulike forventninger til kvalitet og tolkninger om hva som kreves i forhold til læreplaner. Rauner et al. (2013) sitt formål med vurdering er en helhetlig vurdering basert på reelle oppgaver som har en strammere form når det gjelder vurderingskriterier og viktigheten av å trene på bruken av disse.

Når det gjelder *innhold*, er det lite konkrete faglige aspekter å vise til, kun en studie som tar for seg elektrikerfaget (Deichman-Sørensen et al., 2011). Det med el.-sikkerhet ble trukket fram som et svært sentralt tema i forhold til stryk og ikke stryk. Lahn og Nore (2014) viser til at det er stor variasjon når det gjelder hva nemndene la vekt på. Samtidig stiller de spørsmål ved hvordan ulike innholdskomponenter kan bli vurdert. Tsagalidis (2008) viser til at vurdererne som ble fulgt i hennes studie la vekt på både teoretisk og praktisk kunnskap.

Når det gjelder *rammer og kontekstuelle faktorer*, er oppgaver utviklet i nemndene realistiske, men det var stor variasjon i oppgavetyper (Lahn og Nore, 2015). Alt fra åpne komplekse til løsrevne deloppgaver. Rauner (2014) viser til betydningen av å lage reelle oppgaver nært knyttet til arbeidslivet, at disse er en del av et system som er kompetanse- og helhetsorientert. Jacobs (2015) viste til sin studie som pekte på opplevelse av stor relevans i KOMET-oppgavene. I Olsen (2009) vises det til at man må se på nytten og funksjonen til vurderingssystemene og deres intenderte hensikt. I den sammenheng er Friches (2010, 2013) studier viktig. Her pekes på at vurderingspraksiser må tilpasses ulike yrker, og at sosiale og kontekstuelle faktorer spiller inn. Den intenderte hensikten må stadig utvikles.

Når det gjelder *mål og kriterier* viser Deichman-Sørensen et al. (2011) at de karakterer som er gitt viser liten grad av uoverenstemmelse. Spesielt er fokus på stryk karakteren. Vurderingskriterier viser ulik grad av presisering og praksis når det gjelder å informere kandidatene om grunnlaget for vurderingen. Engelsen (1998) har i ett av sine spørsmål vist til behovet for klarhet om hvilke kriterier som skal ligge til grunn. Nore og Lahn (2015) viser til at det i liten grad tas utgangspunkt i kompetansemålene og andre elementer enn de utviklede kriteriene kan trekkes inn. Her nevnes særlig forholdet mellom faglig skjønn og vurderingskriterier. Dessuten utviklingen av den faglige praksisen i vurderingsfellesskapet. Spørsmålet om førsteinntrykk av kandidaten belyses i en vurderingssammenheng. I karaktersetting brukes ofte argumentasjon om man ville ansatt kandidaten, mer enn at man bruker læreplan og kompetansemål. Rafoss og Witsø (2014) viser til behov for større klarhet om kriterier knyttet til karakterer. Vurdererne ønsker i større grad innflytelse på utforming av

kriterier. Rauner et al. (2013) er opptatt av struktur og standardisering av kriterier og løsningsrom. Tsagalidis (2008) viser blant annet til at vurdererne kjente til overordnede mål og kriterier, men i de konkrete situasjonene ble det fokus på hvordan oppgavene ble løst. De studerende blir utsatt for kriterier som er knyttet til yrket mer enn de mål som er satt i skolen. Dette kan være visa versa også at kriterier fra skolen kan overse viktige sider av yrket. Johnson (2008) viste til at vurdererne brukte kriteriene i sin vurdering av portefolier og at dette påvirket deres vurdering.

I omtalen av *metoder* viser Crisp (2012) til hvordan vurderingen gjennomføres. Noen sentrale punkter er kunnskap om studentenes arbeid og planlegging av vurderingsarbeidet. Videre hvordan studenten har realisert oppgaven i forhold til lengde og fokus. Vurderernes emosjonelle og sosiale reaksjoner, fra glede og til det å mislike. Vurdererne kunne også «prate» med studenten, og vurderingen kunne foregå både underveis eller til slutt. Sammenligninger med andre besvarelser forekom. Til slutt var det vesentlig å se over besvarelsen for å gjøre en overordnet og endelig vurdering. Deichman-Sørensen et al. (2011) viser til at det i nemndene var liten grad av uoverenstemmelse mellom medlemmene når det gjaldt å sette en vurdering og karakter.

*Forutsetningenes* betydning viser at det er prøvenemndsmedlemmene selv som skal bedømme den faglige kompetansen (Deichman-Sørensen et al., 2011). De viste til prøvenemndenes utøvelse av skjønn basert på en helhetlig vurdering som bygger på deres egen erfaring, de normer som er etablert og diskusjoner og refleksjoner om den utvikling som ligger i de respektive fagene. Nyen og Tønder Hagen (2014) underbygger dette, og viser til vurderernes normer som er påvirket av arbeidslivets premisser, samtidig som de viser til at læreplanen utgjør en del av vurderingsgrunnlaget. Læreplanen legger et grunnlag for bedømmelser, men faglige normer blant nemndsmedlemmene står ikke eksplisitt uttrykt, og er dermed ikke en del av i læreplanen. Dermed bestemmer nemndsmedlemmenes selv om de har de kunnskaper og ferdigheter som skal til. Når det gjelder å følge loven i vurderingsarbeidet, viser Rafoss og Witsø (2014) til at den ikke følges av den som har lang erfaring som prøvenemndsmedlem. Lahn og Nore (2015) viser til at det helhetlige blikket knyttet til et førsteinntrykk kan påvirke vurderingen. Olsen (2009) viser også til at det gjøres sammenligninger. Rauner et al. (2013) viser til ratere som endret sin rating før og etter trening, hvor konkrete kriterier og rammer utgjorde en vesens forskjell i bedømmingene, og hvor interrater reliabiliteten gikk fra lav til høy. Betydningen av å være to om vurderingene var vesentlig. Tsagalidis (2008) viser til at ulike forutsetninger hos vurdererne (lærere og andre i bransjen) kan føre til at de vurderer



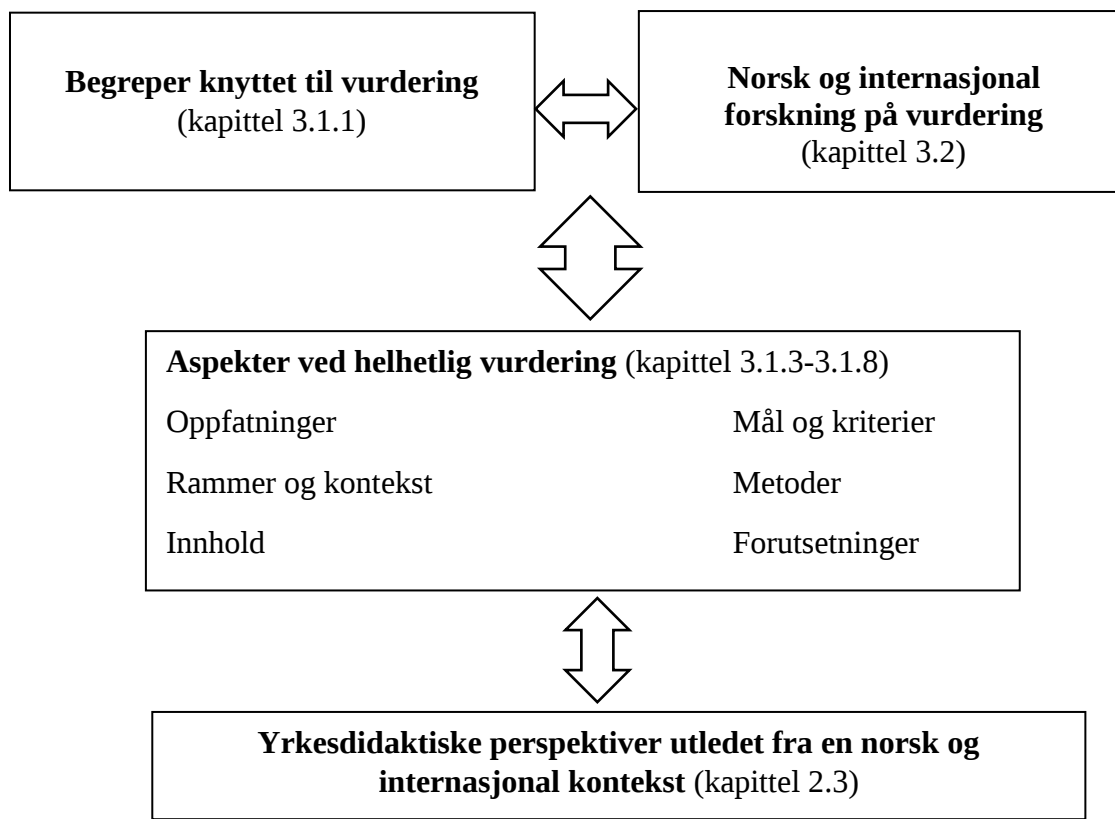
forskjellig. Johnson (2008) bekrefter betydningen av å ha gode relasjoner til den andre vurdereren. Samtidig var det fokus på å gi rom for vurderernes personlige erfaring i vurderingene. Det er også vesentlig eller en forpliktelse å ha kontakt med yrket for å holde seg oppdatert (Johnson 2008).

Videre spørsmål knyttet til ovenstående funn vil være spesielt rettet inn mot bruken av kriterier og skjønn som ligger til grunn for vurderingene. Det vil være sentralt å drøfte dette opp mot teorier som noen av disse studiene ikke gjør, men som kan belyse ytterligere hva som ligger i helhetlig vurdering. Faglig innhold vil være interessant å gå dypere inn i, siden det framkommer lite om det. Videre vil det å forske på flere vurderingssystemer kunne ha betydningen for å se på likheter og ulikheter i vurderingene.

### 3.3 Oppsummering av vurderingsteoretisk rammeverk

Gjennomgangen av utvalgt teori (inkludert yrkesdidaktikk i kapittel 2) utgjør rammeverket for analyse og diskusjon av empiri.

Figur 10 viser sammenhengen mellom de ulike teoretiske delene. Det yrkesdidaktiske feltet omfatter en gjennomgang av en internasjonal og norsk forskning, som har resultert i utviklingen av seks perspektiver om yrkesdidaktikk. Helhetlig vurdering utgjør den andre hoveddelen. Teorien som er presentert er hovedsak knyttet til seks aspekter ved helhetlig vurdering, og som i stor grad bygger på kompetansebasert vurdering. I gjennomgangen ligger redegjørelser for ulike sentrale begreper tilknyttet vurdering. Dessuten er det gitt en gjennomgang av forskning på feltet i internasjonal og norsk kontekst.



Figur 10. En oversikt på teoretisk rammeverk og sammenhengen mellom disse

De teoretiske delene i figuren er gjensidig avhengig av hverandre, og utgjør en helhetlig sammenheng, selv om de er splittet opp som separate deler i framstillingen. Sammenhengen mellom yrkesdidaktikk og helhetlig vurdering vil være sentralt, og vil være til gjenstand for diskusjon, noe som allerede er påbegynt i kapittel 1 innledning og i teorigjennomgangen i kapittel 2 og 3. De sentrale begrepene vil mer eller mindre være med gjennomgående i de kommende kapitlene. Teorier om helhetlig vurdering og forskning presentert i kapittel 3 vil hovedsakelig anvendes i kapittel 5 om presentasjon og diskusjon av resultater og i kapittel 6. De utledede perspektiver om yrkesdidaktikk vil anvendes i kapittel 6 i den sammenfattende diskusjonen.

## 4 Forskningsdesign og metode

I dette kapitlet skal jeg gjøre rede for forskningsdesignet og det metodiske opplegget for innsamling av data i HELVYRD.

Innledningsvis vil jeg ta for meg de tre utvalgte vurderingssystemene. Det vil bli en ytterligere konkretisering som skal gi en forståelse for hva det skal forskes på. Videre vil jeg redegjøre for det overordnede forskningsdesignet for MECVET-prosjektet og vise hvordan HELVYRD er koblet mot dette. Jeg vil beskrive de datasettene som skal gi svar på mine forskningsspørsmål. Deretter tar jeg for meg grunnlaget for intervjuopplegget med en redegjørelse for utvalget av informanter og utvikling av intervjuguider. For hvert vurderingssystem vil jeg ta for meg gjennomføringen av intervjuene. Dokumentanalyse utgjør en mindre del av min datainnsamling, men vil bli viet et mindre delkapittel. Avslutningsvis vil ulike etiske aspekter rundt HELVYRD-undersøkelse bli belyst og kvalitetskriteriene knyttet til dette. Det siste kapitlet setter lys på hvordan analysen er lagt opp og gjennomført for den innsamlede empiri. Det hele avrundes med en diskusjon om kvalitetskriterier innen kvalitativ forskning.

### 4.1 Vurderingssystemene anvendt i HELVYRD-studien

I kapittel 1.3 ble det de norske vurderingssystemene eksamen og fagprøven kort beskrevet. I kapittel 2.2 og 3.1.6 ble sentrale deler av det teoretiske grunnlaget for KOMET-modellen beskrevet. Det dannet grunnlaget for å diskutere tidligere forskning i kapittel 3.2 tilknyttet de tre vurderingssystemene. Her skal de tre vurderingssystemene beskrives mer konkret.

Beskrivelsen har følgende struktur: (1) *Vurderingssystem og gjennomføring*, (2) *Oppgavetyper* og (3) *Vurderingsprosessen*.

#### 4.1.1 KOMET-modellen anvendt i MECVET

Gjennomgangen av MECVET-testen vil vise hvordan KOMET-modellen er anvendt og tilpasset i den norske konteksten.

##### *Vurderingssystem og gjennomføring*

I MECVET-prosjektet er hoveddrammene for gjennomføring av testene fulgt. Det har foregått ved at forskere i prosjektet har reist til de respektive skoler og opplæringskontorer og møtt lærere/instruktører og elever/lærlinger. Informasjon har vært gitt og testene gjennomført etter bestemte rutiner med fordeling av oppgaver og elev/lærlingene har hatt to timer på seg til å besvare. Forskjellene ved MECVET er at hver elev/lærling har hatt full tilgang til alle

hjelpemidler inkludert internett, og at det skal brukes PC for å skrive besvarelsen.

(Gjennomføring av testen i Tyskland innebærer en 2 timer skriftlig test med penn og papir og de har ikke lov til å bruke hjelpemidler.) Alle forskerne går rundt og samler inn besvarelsene på en minnebrikke og tar de med seg for rating.

### Oppgavetyper

Når det gjelder testoppgavene i MECVET, bygger disse på prinsippene fra KOMET-modellen. Oppgavene utviklet for elektriker-faget i Tyskland passet ikke i Norge. De norske ekspertene har utviklet nye oppgaver basert på rammen av KOMET-tenkningen, sett i en norsk kontekst. De fire oppgavene som er utviklet har blitt laget for å gi et bilde av hva faget innebærer, hva elever og lærlinger kan forvente seg i yrket. Figur 11 viser et eksempel på oppgave D, som handler om en oppgradering av en familiehytte. Den er delt inn i tre hovedområder.

<p><b>Oppgave D</b></p> <p><b>Oppgradering av familien Pettersens hytte</b></p> <p><b>Situasjonsbeskrivelse</b></p> <p>Familien Pettersen eier en hytte på Blåøy. Den ble bygget på 1960-tallet. Nå ønsker de å oppgradere og pusse opp samt etterisolere ytterveggene.</p> <p>Hytta har et gammelt åpent elektrisk anlegg og som en del av oppussingen ønsker familien å oppgradere det elektriske anlegget. Mye av det elektriske anlegget som stikkontakter og ledninger er demontert, slik at det må legges opp nytt. Familien Pettersen ønsker å se på muligheten for noe gjenbruk av eksisterende anlegg og utstyr. De ønsker videre å bytte ut varmekildene (gamle panelovner) i hytta til noe som er mer økonomisk og miljømessig riktig.</p> <p>Hytta har ikke hatt innlagt vann så Pettersen har kjøpt inn en pumpe type B 20 (se vedlegg), som han skal benytte til å pumpe vann opp fra et borehull og inn i hytta. Han vil ha en løsning for å styre pumpa fra hytta.</p> <p>Inntaket til hytta er luftstrekk til inntaksskap på hytteveggen.</p>	<p><b>Oppgavetekst</b></p> <p>Pettersen ber om en vurdering av eksisterende anlegg og utstyr og forslag til nytt elektrisk anlegg inkludert nye varmekilder og styring for pumping av vann.</p> <p>Installatøren har bedt deg om å utarbeide et løsningsforslag som utgangspunkt for at firmaet kan få levert familien Pettersen et anbud. Dersom du trenger å avklare noen spørsmål underveis – for eksempel med kunden, leverandører eller andre håndverkere – ta dette med i beskrivelsen.</p> <p>Begrunn løsningsforslaget.</p> <p><b>Verktøy og hjelpemidler</b></p> <p>Alle hjelpemidler tillatt.</p> <p><b>Vedlegg:</b></p> <ol style="list-style-type: none"><li>1. Tekniske data på forsyningsanlegget</li><li>2. Skissetegning eksisterende anlegg</li><li>3. Produktdatablad, pumpe</li><li>4. Vurderingskriterier for testoppgaven</li></ol>
--	--

Figur 11. Situasjonsbeskrivelsen og oppgavetekst (MECVET etter modell fra KOMET (Rauner, Haasler, Heinemann & Grollmann, 2009, s. 184)

Første del er en situasjonsbeskrivelse av arbeidsoppgaven. Den andre delen er oppgaveteksten, og den tredje og siste delen er informasjon om verktøy og hjelpemidler. I tillegg til dette kommer det vedlegg. Situasjonsbeskrivelsene skal sette kandidaten inn i problemet. Teksten er konkret og skal gi den nødvendige informasjonen og beskrive ulike sider ved situasjonen. Videre kommer selve oppgaveteksten som forklarer hva kandidaten skal gjennomføre. Teksten er kort og konkret, men samtidig åpen. Den tredje delen er en kort informasjon om hvilke verktøy og hjelpemidler som kan benyttes. I disse oppgavene er alle

hjelpemidler tillatt. De tre første vedleggene inneholder konkret nødvendig teknisk informasjon pluss skisser og tegninger. Bakerst i oppgaven ligger det vurderingskriterier i form av de åtte dimensjonene samt et løsningsrom. Jeg tar for meg dette i neste avsnitt.

### Vurderingsprosessen

Når det gjelder vurderingsprosessen, har raterne vært igjennom en opplæring i hvordan rating gjennomføres i lys av KOMET-modellens tenkning. Jeg vil utdype dette noe mer, siden dette er fra en modell som er utenfor det norske systemet. Treningen foregår over to dager hvor man blant annet bruker utplukkende besvarelser som grunnlag. Målet for raterne ligger i det å utvikle en forståelse for metoden som benyttes, samt skape en omforent rating av de enkelte besvarelser (rater parvis). Når det gjelder til ratingen, er det utviklet både et ratingskjema og et løsningsrom tilpasset norske forhold. Til hver besvarelse skal ett ratingskjema benyttes. Se figur M2.

Vurderingsskjema elektrikerfaget		Kriterier:				Kriteriene er...			
Kandidat:		ikke	delvis	nesten	ikke	delvis	nesten	opppt	opppt
Navn:		opppt	opppt	opppt	opppt	opppt	opppt	opppt	opppt
Oppgave:									
<b>(1) Presentasjon av innhold</b>									
Er beskrivelsen klar nok til å kunne legges frem for en arbeidsgiver/oppdragsgiver?									
Er beskrivelsen forståelig for fagfag?									
Er løsningen tydelig presentert (skriftlig, grafisk)?									
Er finansdelen strukturert og oversiktlig?									
Passer beskrivelsen til oppdraget?									
<b>(2) Faglig funksjonalitet</b>									
Følger løsningen?									
Følger forslaget gjeldende forskrifter, standarder og praksis i faget?									
Vurderes det om løsningen er praktisk gjennomførbar?									
Er løsningen faglig fornuftig og godt begrunnet?									
Er beskrivelsen og forklaringene riktige?									
<b>(3) Bruksevne/holdbarhet</b>									
Er løsningen enkel å vedlikeholde og reparere?									
Står det tett hensyn til langiktig bruk og muligheter for utvidelser eller fornyelser?									
Vurderes og begrunnes eventuelle svake punkter og mangler?									
Hvite beskrivende og forklarende for den som skal utføre arbeidet?									
Hvite beskrivende og forklarende for oppdragsgiver/handen?									
<b>(4) Effektivitet/økonomi</b>									
Er løsningen vurdert i forhold til installasjonskostnader og lønnsomhet?									
Tar medtelling av løsningen hensyn til personressurser og tidbruk?									
Vurderes ressurseffektivitet og forbruksmønstre opp mot lønnsomhet?									
Er utførelses- og vedlikeholdskostnader vurdert?									
Inneholder løsningen en effektiviseringsgevinst for bedriften eller kunden?									
<b>(5) Organisering av arbeidet</b>									
Er løsningen forankret i bedriftens arbeidsprosesser og organisasjon?									
Hvite løsningen knyttet til arbeidsprosesser?									
Tar løsningen hensyn til utførelsen av oppgaver i arbeidsprosessen?									
Tilbyr løsningen ytterligere muligheter i arbeidsprosessen?									
Betrakter løsningen samarbeid og oppgaver med tilgrensede fag?									
<b>(6) Arbeidsmiljø/ HMS</b>									
Er løsningen hensyn til trykksikkerhet arbeidsmiljø?									
Står spørsmål om helse/sikkerhet tett opp og vurdert?									
Er ergonomi vurdert?									
Følger løsningen gjeldende retningslinjer for HMS?									
Vurderes sosiale og samfunnsrettede konsekvenser?									
<b>(7) Miljøsvær</b>									
Tar forslaget hensyn til gjeldende miljøbestemmelser i samfunnet?									
Står det tett hensyn til miljøsværhet i løsningen?									
Hvite grad tar løsningen hensyn til miljøsværhet arbeidsorganisering og produksjon (green building)?									
Tar løsningen opp spørsmål rundt gjennomgang, gjennom og miljøsværhet bærekraft?									
Tar forslaget opp spørsmål om energisparing og energieffektivisering?									
<b>(8) Kreativitet</b>									
Inneholder beskrivelsen elementer til nye forutsetninger?									
Står ulike kvalifikasjoner vurdert opp mot løsningen?									
Er utførelsen av løsningen utvetydig kvaliteter ifrå estetikk?									
Er løsningen utvetydig for god problemforståelse?									
Står handlingsrommet i oppgaven godt utnyttet?									

Figur 12. Vurderingsskjema med åtte dimensjoner og 40 items (MECVET etter modell fra KOMET (Rauner et al., 2009, s. 182-183))

Ratingskjemaet inneholder åtte MECVET-kriterier (dimensjoner) med tilhørende 5 kriteriespesifikasjoner (veiledende). MECVET-kriteriene har følgende betydning (oversatt med min tolkning).

1. Det første kriteriet omhandler *funksjonaliteten* i den tekniske løsningen. Den viser til den tekniske kompetansen, eller til kontekst-uavhengig subjektspesifikk kunnskap og ferdighet. Beviset på at funksjonaliteten er til stede, er vesentlig for de øvrige kravene til løsning for de enkelte arbeidsoppgavene.

2. *Presentasjon* er resultatet av at jobben skal dokumenteres og presenteres. Dette er nødvendig for at de som skal ha den foreslåtte løsningen lett skal kunne forstå og revidere det

som blir lagt frem. Et kjernepunkt i kommunikasjonen av en arbeidskontekst er å kunne uttrykke sine tanker på en klar og organisert måte gjennom tegninger og skisser av løsningen.

3. *Effektivitet* er en nødvendighet i profesjonelt arbeid for å evaluere det økonomiske aspektet ved en oppgave, blant annet forutse ulike kostnader som kan påvirke arbeidet.

4. *Holdbarhet* omhandler de arbeidsprosesser og arbeidsoppgaver som utføres, og er rettet inn mot kundens behov. I tillegg til at løsningen skal fungere for kunden umiddelbart, er det viktig å være proaktiv med tanke på at feil kan oppstå, og at det skal være lett å vedlikeholde og reparere.

5. *Arbeidsprosessorientering* tar utgangspunkt i forståelsen av organisasjonens hierarki og kjeden av prosesser. I tillegg skal det tas hensyn til foranliggende og etterfølgende arbeidsprosesser som også inkluderer aspekt av samarbeid som er på tvers av eget arbeid.

6. *Miljøansvar* knyttes til de konkrete arbeidsoppgaver og arbeidsprosesser, at disse er i henhold til de krav som stilles. Det kan være materialer som brukes eller energisparetiltak og resirkulering.

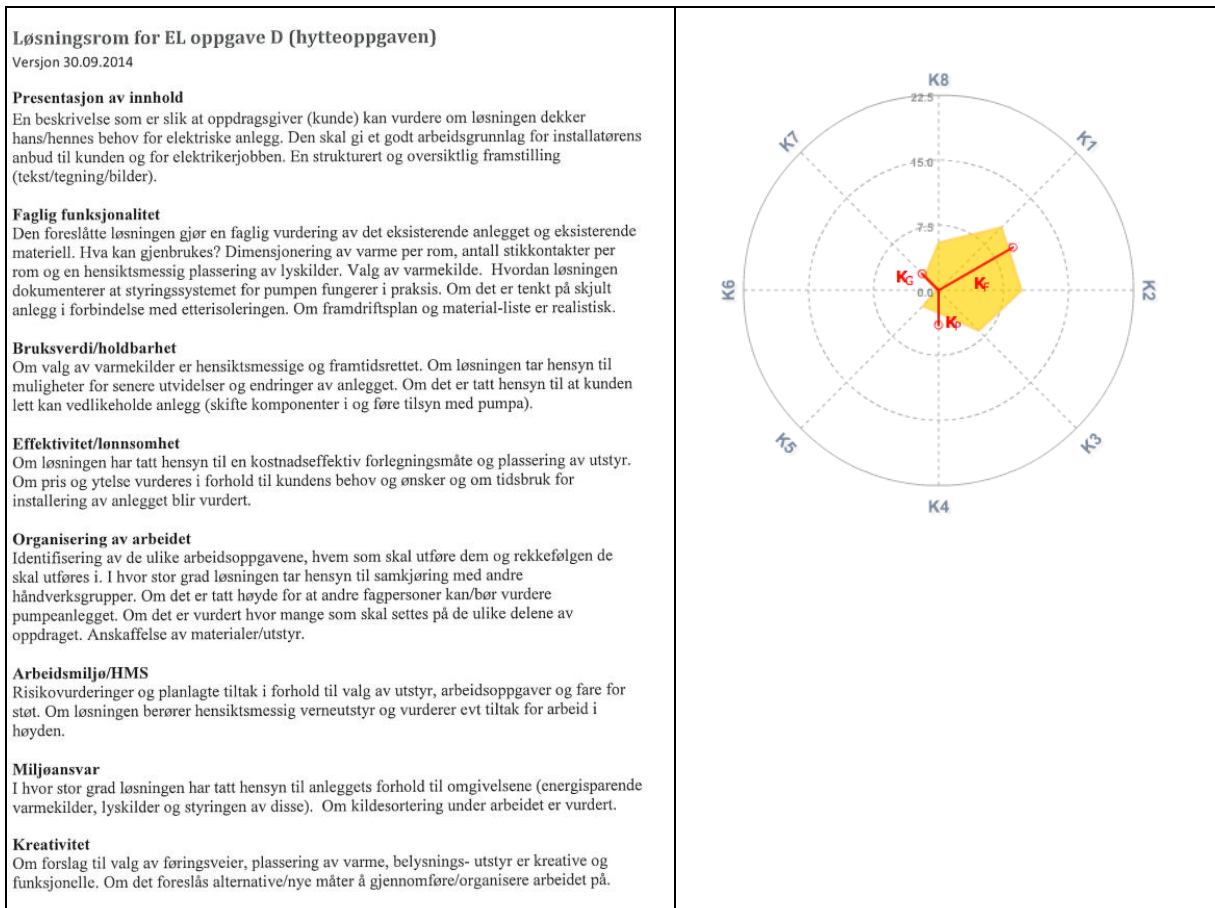
7. *Kreativitet* i de løsninger som er forelagt forventes av de kompetente utøverne i yrkene, slik at uvanlige løsninger oppleves som meningsfulle bidrag i den endelige løsningen.

8. *Arbeidsmiljø/HMS* kriteriet referere til organiseringen av arbeidet med fokus på menneskene samt helse og det sosiale aspektet. Det inkluderer sikkerhet på arbeidsplass for å hindre ulykker, og at de løsninger som lages tar hensyn til det miljøet de skal fungere i.

Hver av de 40 (8x5) kriteriespesifikasjonene tilknyttet en ratingskala på fire nivåer (ikke oppfylt – dårlig oppfylt – nesten oppfylt – oppfylt) med respektivt tilhørende tall 0-3.

Resultatene av ratingen fra den enkelte rater blir plottet inn i et skjema og det blir gjort beregninger. Skåren på de enkelte besvarelser skal drøftes av minst to ratere for å oppnå konsensus.

For hver oppgave er det utviklet et løsningsrom (figur13, venstre). Hensikten er å gi noen indikasjoner på hva som forventes i besvarelsen som veiledning for raterne uten å være noe fasitsvar. Her er det forslag knyttet til hver enkelt dimensjon.



Figur 13. Løsningsrom (venstre) og eksempel på kompetanseprofil (høyre) (MECVET etter modell fra KOMET (Rauner et al., 2009, s. 185))

Sluttresultatet blir presentert i en kompetanseprofil (spindel) (figur 13, høyre), hvor de åtte MECVET-kriteriene er vist (et eksempel).

#### 4.1.2 Sentralgitt skriftlig eksamen

Når det kommer til *sentralgitt skriftlig eksamen* er det en oppgave som utvikles sentralt, gjennomføres nasjonalt og sensureres lokalt.<sup>1</sup>

#### Vurderingssystem og gjennomføring

Innenfor yrkesfagene gjennomføres det sentralgitte skriftlige eksamener for elever, lærlinger og praksiskandidater i enkelte utvalgte fag før man går opp til fagprøven. Dette gjelder innenfor elektrikerfaget (og blir gjennomført ca halvt år for fagprøven). Det er en skriftlig eksamen inkludert en forberedelsesdel og selve gjennomføringen skal følge de retningslinjer som er gitt i Forskrift til opplæringslova (2006, §3-52 til §3-55) og veiledningen fra

<sup>1</sup> Det har vært gjort endringer og justeringer i forskriften (november 2018) siden datainnsamlingen ble gjennomført våren 2015. Det har ikke nevneverdig betydning for min studie, men henvisningene til forskriften er til den nyeste versjonen. Det gjelder også for fagprøven i kapittel 4.1.3.

Utdanningsdirektoratet (2018). Kandidatene (lærlinger Vg3) sitter spredt rundt omkring i Norge, og oppgavene blir distribuert gjennom Fylkeskommunen. Rammene for gjennomføring innebærer en 5 timers sentralgitt skriftlig eksamen med en forberedelsesdel som starter 24 timer før selve eksamen.

### *Oppgavetyper*

Ved utarbeiding av eksamensoppgaven skal det tas utgangspunkt i læreplanene for Vg3 i lærefaget, og de kompetansemål som står beskrevet der. Kompetansemålene skal uttrykke den forventning man har til hva kandidaten skal vise i sin besvarelse. Når det gjelder eksamensoppgaven, uttrykkes det i veiledningen at *«Nokre oppgåver vil vere svært opne og overlata til kandidaten å gjere mange val, mens andre kan ha fleire konkrete "bestillingar" i form av deloppgåver, spørsmål eller krav om at kandidaten må ha med bestemte moment.»* (Utdanningsdirektoratet, 2018, s. 3). Dette viser til det mangfold av ulike type eksamensoppgaver som kan utformes og møte fagenes særegenhet.

Figur 14 (venstre) viser informasjon om forberedelsesdelen til lærlingene. Den varer én dag, og gir informasjon om hjelpemidler, kildebruk, diverse opplysninger og informasjon om vurdering. Videre får lærlingene utlevert en tema/situasjonsbeskrivelse (Figur 14, høyre). I dette tilfellet handler det om å planlegge, gjennomføre og dokumentere to ulike arbeidsoppdrag. Ett arbeidsoppdrag handler om en rehabilitering av en bolig (nytt kjøkken), og det andre er knyttet til en leilighet med utbedring av seks ulike tekniske punkter. Det benyttes et standardskjema som knyttes til begge arbeidsoppdragene. Der står det informasjon om kunden, hva ordren går ut på, samt en litt mer utfyllende ordrebekrivelse. Dette er alt som kandidatene får utlevert til denne dagen utenom den formelle informasjonen.



Bokmål		Tema/Situasjonsbeskrivelse	
<b>Informasjon til forberedelsesdelen</b>		Du er ved slutten av læretiden din og læringens ansvarlig har bestemt at du på egenhånd skal planlegge, gjennomføre og dokumentere to arbeidsoppgaver. Oppdragene er av forskjellig art. Det ene er rehabilitering av enebolig og det andre er utbedring av punkter i forbindelse med kontroll i en leilighet.	
Forberedelsestid	Forberedelsestiden varer én dag.	KUNDE: Gro Glad	ORDRE NR: 01 V/2015
Hjelpemidler	På forberedelsesdagen er alle hjelpemidler tillatt, inkludert bruk av Internett.  På eksamen er alle hjelpemidler tillatt, bortsett fra Internett og andre verktøy som kan brukes til kommunikasjon.	TLF. 44332211	KONTAKTPERSON: Gro Glad
Bruk av kilder	Hvis du bruker kilder i besvarelsen din, skal disse alltid oppgis på en slik måte at leseren kan finne fram til dem.	Utstyrstype:	Merke og type: Serienummer:
Andre opplysninger	I forberedelsestiden kan du samarbeide med andre, finne informasjon og få veiledning.	Ordre:	• Rehabilitering av enebolig.
Informasjon om vurderingen	Jf. forskrift til opplæringslova §§ 3-52, 3-53, 3-54 og 3-55: "Eksamen skal prøve i kva grad den enkelte praksiskandidaten/lærlingen har nådd kompetansemåla som er fastsette i den læreplanen (Vg3) for opplæring i bedrift som praksiskandidaten/lærlingen tek sikte på å ta fag- eller sveineprøve/har teikna lærekontrakt i."	Mottatt av: Anders Watt, Saksbehandler	Dato: 29. mai 2015
		Ordrebeskrivelse:	• Det skal monteres nytt kjøkken og kunden ønsker seg nytt elektrisk anlegg i tilknytning til dette
		Måledata:	
		Medgått tid (timer):	
		Ferdig dato:	Monter:
		KUNDE: Peder Pedersen	ORDRE NR: 2 V/2015
		TLF. 66473829	KONTAKTPERSON: Peder Pedersen
		Utstyrstype:	Merke og type: Serienummer:
		Ordre:	• Utbedring av feil, påpekt ved kontroll/tilsyn
		Mottatt av: Anders Watt, Saksbehandler	Dato: 22. mai 2015
		Ordrebeskrivelse:	• Dårlig eller ingen hovedjord • Varmgang i tilførselsledning til varmtvannsbøder • Potensialforskjell målt mellom vannledning og vaskemaskin • Varmgang sikringskapp • Sprekk i flere sikringsholdere til uz elementer • Mange skjoteledninger tilkoblet stikkontakt bak TV
		Måledata:	
		Medgått tid (timer):	
		Ferdig dato:	Monter:

Figur 14. Informasjon til gjennomføring av forberedelsesdelen (venstre), og beskrivelse av situasjonen (høyre). (Kilde udir.no)

Til selve eksamensoppgaven gis det informasjon (Figur 15 venstre) om eksamens varighet og hvilke vedlegg som er lagt ved.

Bokmål		Utfyllende informasjon til ordre nr. 01V / 2015.		Vedlegg 1												
<b>Eksamensinformasjon</b>		<ul style="list-style-type: none"> <li>Rommet der kjøkkenet skal være er klart for elektriker. Alle plater er fjernet og nye vegger er satt opp. Alle vegger er isolert. Se vedlegg 1.</li> <li>Kunden ønsker følgende kjøkkenmaskiner: <ul style="list-style-type: none"> <li>Induktiv platetopp</li> <li>Støkeovn</li> <li>Oppvaskmaskin</li> <li>Mikrobølgeovn</li> <li>Integrert kaffetrakter</li> <li>Kjøleskap</li> <li>Avtrekksvifte</li> </ul> </li> <li>Det nye kjøkkenet er utvidet etter at det har blitt slått sammen med et tidligere soverom. Kunde ønsker varme i gulvet. Ferdig gulv skal være parkett.</li> </ul>		<p>Vedlegg 1</p> <p>Vedlegget skal leveres inn sammen med det andre svarark. / Vedlegget skal leveres inn sammen med de andre svararkene.</p> <table border="1"> <thead> <tr> <th>Nr. til:</th> <th>Salt/Saltet på svarark</th> <th>Narmer/narmer på skolen</th> <th>Disseminert til:</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>		Nr. til:	Salt/Saltet på svarark	Narmer/narmer på skolen	Disseminert til:							
Nr. til:	Salt/Saltet på svarark	Narmer/narmer på skolen	Disseminert til:													
<table border="1"> <tr> <td>Eksamenstid</td> <td>Eksamen varer i 5 timer.</td> </tr> <tr> <td>Hjelpemidler</td> <td>Alle hjelpemidler er tillatt, bortsett fra Internett og andre verktøy som kan brukes til kommunikasjon.</td> </tr> <tr> <td>Bruk av kilder</td> <td>Hvis du bruker kilder i besvarelsen din, skal disse alltid oppgis på en slik måte at leseren kan finne fram til dem.</td> </tr> <tr> <td>Vedlegg</td> <td>1. Plantegning av kjøkken</td> </tr> <tr> <td>Vedlegg som skal leveres inn</td> <td>1. Plantegning av kjøkken</td> </tr> <tr> <td>Informasjon om vurderingen</td> <td>Jf. forskrift til opplæringslova §§ 3-52, 3-53, 3-54 og 3-55: "Eksamen skal prøve i kva grad den enkelte praksiskandidaten/lærlingen har nådd kompetansemåla som er fastsette i den læreplanen (Vg3) for opplæring i bedrift som praksiskandidaten/lærlingen tek sikte på å ta fag- eller sveineprøve/har teikna lærekontrakt i."</td> </tr> </table>	Eksamenstid	Eksamen varer i 5 timer.	Hjelpemidler	Alle hjelpemidler er tillatt, bortsett fra Internett og andre verktøy som kan brukes til kommunikasjon.	Bruk av kilder	Hvis du bruker kilder i besvarelsen din, skal disse alltid oppgis på en slik måte at leseren kan finne fram til dem.	Vedlegg	1. Plantegning av kjøkken	Vedlegg som skal leveres inn	1. Plantegning av kjøkken	Informasjon om vurderingen	Jf. forskrift til opplæringslova §§ 3-52, 3-53, 3-54 og 3-55: "Eksamen skal prøve i kva grad den enkelte praksiskandidaten/lærlingen har nådd kompetansemåla som er fastsette i den læreplanen (Vg3) for opplæring i bedrift som praksiskandidaten/lærlingen tek sikte på å ta fag- eller sveineprøve/har teikna lærekontrakt i."	<p><b>Utfyllende informasjon til ordre nr. 2V / 2015.</b></p> <ul style="list-style-type: none"> <li>Kontroller har målt overgangsmotstand til jord på 3200Ω. Kunde kan bekrefte at vannledning ble byttet ut for noen år siden. Kobber ble erstattet med plast. Vanntilførsel ble tidligere benyttet som jordelektrode, tilkoblinger er intakt.</li> <li>Varmtvannsbøder er ca. 8 år gammel og er tilkoblet via en stikkontakt. Ved pluggen kan det tydelig høres at ledningen er varm.</li> <li>Kunde kan bekrefte potensialforskjell mellom vaskemaskin og vannrør, da han flere ganger har fått støt.</li> <li>Kurs 3 og 5 har betydelig varmgang både i deksler og ledninger.</li> <li>Kunden ønsker informasjon om hva som må gjøres for å oppgradere det elektriske anlegget i samsvar med dagens krav.</li> </ul>			
Eksamenstid	Eksamen varer i 5 timer.															
Hjelpemidler	Alle hjelpemidler er tillatt, bortsett fra Internett og andre verktøy som kan brukes til kommunikasjon.															
Bruk av kilder	Hvis du bruker kilder i besvarelsen din, skal disse alltid oppgis på en slik måte at leseren kan finne fram til dem.															
Vedlegg	1. Plantegning av kjøkken															
Vedlegg som skal leveres inn	1. Plantegning av kjøkken															
Informasjon om vurderingen	Jf. forskrift til opplæringslova §§ 3-52, 3-53, 3-54 og 3-55: "Eksamen skal prøve i kva grad den enkelte praksiskandidaten/lærlingen har nådd kompetansemåla som er fastsette i den læreplanen (Vg3) for opplæring i bedrift som praksiskandidaten/lærlingen tek sikte på å ta fag- eller sveineprøve/har teikna lærekontrakt i."															
	<b>Planlegg, gjennomfør og dokumenter oppdragene.</b>															

Figur 15. Informasjon til gjennomføring eksamen (venstre), utfyllende informasjon til de enkelte ordrene (midten) og skisse over plantegning (høyre). (Kilde Udir.no)

Videre gis det mer utfyllende informasjon om selve eksamensoppgaven knyttet til hver arbeidsordre (Figur 15, midten). Dette handler om ren teknisk informasjon, informasjon om

utstyr og aktuelle rom hvor jobben skal gjøres. Eksamensoppgaven avsluttes med «Planlegg, gjennomfør og dokumenter oppdragene». Det er også lagt ved et vedlegg som viser en plantegning (Figur 19, høyre).

### *Vurderingsprosessen*

Når det gjelder vurderingen er det utnevnt sensorer i hver enkelt fylkeskommune. I grunnlaget for vurderingen står det beskrevet:

Sensorane skal vurdere i kva grad kandidaten viser kompetanse slik han er beskriven i kompetansemåla i læreplanen. Sensorane har derfor ikkje høve til å stille krav som går ut over innhaldet i læreplanen. Sensorane skal vurdere korleis innhaldet i svarea er frå eksamensoppgåva. (Utdanningsdirektoratet, 2018, s. 4)

Sensor skal ikke trekke inn elementer som går utover det som står beskrevet som kompetansemål i læreplanen, og at selve oppgaven knyttet til de aktuelle kompetansemål skal utgjøre grunnlaget for det som blir vurdert.

En karakter skal gis på individuelt grunnlag basert på den kompetansen som kommer fram på eksamen, og forberedeledelsen skal ikke vurderes.

Karakterskalaen går fra 1 til 6 (1 er ikke bestått) og det vises til Forskrift til opplæringslova (2006, §3-4 tredje ledd) for generelle beskrivelser av de ulike karakterene i grunnopplæringa.

Disse gjengis under:

- a) Karakteren 6 uttrykkjer at eleven har framifrå kompetanse i faget.
- b) Karakteren 5 uttrykkjer at eleven har mykje god kompetanse i faget.
- c) Karakteren 4 uttrykkjer at eleven har god kompetanse i faget.
- d) Karakteren 3 uttrykkjer at eleven har nokså god kompetanse i faget.
- e) Karakteren 2 uttrykkjer at eleven har låg kompetanse i faget.
- f) Karakteren 1 uttrykkjer at eleven har svært låg kompetanse i faget.

(Utdanningsdirektoratet, 2018, s. 4)

Ved selve besvarelsen for kandidatene er det uttrykt følgende i veiledningen:

Når kandidaten vel å bruke erfaringar frå eigen praksis, er det viktig at ho/han beskriv klart kva for val som blir tekne, og dei føresetnadene det blir arbeidd ut ifrå.

Kandidaten må halde seg til det oppgåva spør etter og bruke fagkompetansen sin.

(Utdanningsdirektoratet, 2018, s. 3)

Det vises til åpenhet omkring det å bruke egne erfaringer så lenge man beskriver hvordan og på hvilke forutsetninger dette skjer, og at man holder seg til oppgaveteksten.

Når det gjelder vurderingsprosessen for de konkrete sensorene i denne studien, blir eksamensoppgaven og besvarelsene sendt til de respektive. Sensorene jobber parvis, og det enkelte par får de samme besvarelsene. Først sitter de enkeltvis hver for seg og sensurer, og møtes deretter for samsensur. For samsensur møttes alle i denne sensurgruppa til en diskusjon rundt eksamensoppgaven. Hensikten er å samkjøre, samt å sette fokus på det de mener oppgaven handler om. Etter diskusjonen går de i sine parvise sensorgrupper og gjennomfører samsensur og blir enige om de karakterene den enkelte besvarelse skal få. Det blir ikke registrert noe skjema eller andre dokumenter som viser til at det er utviklet konkrete vurderingskriterier som blir benyttet ved sensurering.

#### 4.1.3 Fagprøven

Fagprøven er den avsluttende vurdering for å få et fagbrev og som dokumenterer at lærlingen har oppnådd en kompetanse som fagarbeider innen sitt yrke.

##### *Vurderingssystem og gjennomføring*

Fagprøven markerer avslutningen på et utdanningsløp (2+2 normalløp, elektriker har avvik med 2,5 år læretid) (Udir Utdanningsløp). I Forskrift til opplæringslova (2006, §3-48) står det at fagprøven skal gjennomføres ved slutten av opplæringstida, og normalt i den bedriften lærlingen har tatt hoveddelen av sin opplæring. Det er også mulighet for å holde prøve et annet sted etter råd fra prøvenemnda. I Forskrift til opplæringslova (2006, §3-52) står det at fylkeskommunen oppnevner prøvenemndene. Disse skal bestå av minst to medlemmer med formell faglig kompetanse. Oppdatert arbeidslivserfaring skal etterstrebes så langt det er mulig. Det presiseres også at den vurderingsfaglige kompetanse må være tilfredsstillende.

Fagprøvene i min studie gikk over åtte dager med oppstart på onsdag i den ene uken, og sluttvurdering på en torsdag påfølgende uke. Fagprøvene som er valgt for HELVYRD-prosjektet ble gjennomført på en prøvestasjon. Det vil si at dette foregår i et lokale hvor det var bygd opp flere båser. Lærlingene ble plassert i hver sin bås med sine respektive oppgaver. Dette ble gjennomført parallelt og hver prøvenemnd følger fire lærlinger i samme uke. Første delen av fagprøven er planleggingsdel (onsdag, torsdag og fredag med fremlegg av dette til prøvenemnda på mandag). Den neste delen er gjennomføring (mandag, tirsdag og onsdag). Siste delen er torsdag, med sluttvurdering og samtaler med lærlingen. Her blir det gjort en

gjennomgang av all dokumentasjon og egenvurdering. Prøvenemnda tar for seg en og en lærling denne dagen (fire lærlinger). I løpet av denne uka er prøvenemnda innom flere ganger underveis og følger opp lærlingene. Gjennomføringen er utformet i henhold til de retningslinjer som er gitt i Forskrift til opplæringslova (2006, §3-54).

### *Oppgavetyper*

Når det gjelder oppgavens konstruksjon, står det i Forskrift til opplæringslova (2006, §3-53) at:

Fag- og sveineprøva skal prøve kandidaten sin kompetanse i faget etter læreplanen for opplæring i bedrift. Alle kompetansemåla i læreplanen for faget skal kunne prøvast. Prøvenemnda skal syte for at kandidaten får vist om han eller ho har den kompetansen som skal til for å få fag- eller sveinebrev i faget.

Fag-/sveineprøva skal prøve kandidaten i å planlegge, gjennomføre, vurdere og dokumentere eit eige fagleg arbeid.

(Forskrift til opplæringslova, 2006, §3-53)

Det er viktig at prøvenemnda utformer omfanget av prøven slik at kandidaten får mulighet til å få vist sin kompetanse innenfor den gitte tidsrammen, og det er også mulig å ha en faglig samtale til slutt for avklaringer. Hjelpemidler kandidaten er vant med å bruke i sin læretid skal også være tilgjengelige under fagprøven. Prøvenemnda skal også utvikle et grunnlag som skal si noe om hva som skal vurderes, og dette skal på gjøres kjent for kandidaten (Forskrift til opplæringslova, 2006, §3-53).

Når det gjelder selve de konkrete oppgavene i denne studien er det utviklet flere sett. Jeg beskriver oppbyggingen av disse prinsipielt og viser ikke konkrete bilder grunnet konfidensialitet. Oppgavene er utformet slik at de skal passe til de tre veggene i den utleverte båsen. Man kaller det hovedvegg, høyrevegg og venstrevegg. Til hver av disse veggene er det utformet en konkret tekst med tekniske opplysninger og instruksjon om hva som skal gjøres av arbeidsoppgaver. Noen oppgaver har også med litt mer tekst om gjennomføring som også gir en instruksjon om hva som skal gjøres. Det følger med en rekke vedlegg som teknisk informasjon, skisser, tegninger, skjemaer for beregninger, skjemaer for ulike kontroller, utstyrslistene og produktinformasjon. Vurderingskriterier er også vedlagt.

### *Vurderingsprosessen*

Ved gjennomføringen av fagprøven skal prøvenemnda være til stede ved prøvestart og prøveslutt. Det skal føres tilsyn underveis og prøvenemnda skal selv vurdere kandidaten sitt arbeid (Forskrift til opplæringslova, 2006, §3-54).

Ved selve vurderingen av kandidatens fagprøve står det uttrykt i Forskrift til opplæringslova (2006, §3-55) at det skal skje gjennom den kompetanse som er dokumentert.

Det skal settes en karakter som skal gi uttrykk for i hvilken grad målene i læreplanen er nådd, og denne tredelte skalaen viser til Forskrift til opplæringslova (2006, §3-4 sjette ledd). Denne skalaen er som følger:

- a) Bestått mykje godt: mykje god eller framifrå kompetanse i faget
- b) Bestått: tilfredsstillande eller god kompetanse i faget
- c) Ikkje bestått: svært låg kompetanse i faget

Skulle det komme til uenighet i karaktersettingen, er det fylkeskommunen som på forhånd har bestemt hvem skal avgjøre hvilken karakter det blir. Der vil vedkommende ha en dobbeltstemme.

På siste dag har prøvenemnda en gjennomgang av den praktiske løsningen i båsen sammen med en lærling. Etter det tar nemnda med seg lærlingen til et rom hvor det blir en gjennomgang av dokumentasjon og en egenvurdering. Deretter skal nemnda sitte alene og bestemme seg for en karakter og deretter blir lærlingen informert om resultatet. Dette er den normale prosedyren.

Prøvenemndene skal vurdere med henblikk på kompetansemålene og ved denne prøvestasjonen er det utviklet et skjema med overskrift «Vurderingskriterier for fagprøve elektriker». Lærlingene ble informert om dette og kriteriene er vedlagt oppgaven. Skjemaet for vurderingskriteriene er utformet i henhold til de fire delene en fagprøve består av. Jeg beskriver også disse kort prinsippielt. Ved *planlegging* er det syv områder som utgjør kriteriene. Det er beregninger, ulike skjemaer, fremdriftsplan, sikkerhet, materiell, ytre påvirkninger og jording. Når det gjelder *gjennomføring* er det åtte områder hvor det under hver av disse spesifiserte underpunkter. Områdene er fordelingsanlegg/sikringssskap, jording, lysanlegg/varmeanlegg/kursopplegg, avfallshåndtering, svakstrømsanlegg, asynkronmotoranlegg, forskrifter/normer og feilsøking/feilretting. Ved *dokumentasjon* er det fire hovedområder med spesifiserte underpunkter. Områdene er forskrifter/norm,

egenkontroll-/sluttkontrollskjema, helse, miljø og sikkerhet, dokumentasjon. Ved *egenvurdering* er det ett område kalt egenvurdering med begrunnelse.

Kriteriene skal utgjøre en ramme for den enkelte nemndsmedlem i deres arbeid med å vurdere lærlingenes besvarelser.

#### 4.2 Forholdet MECVET og HELVYRD

I innledningen ble det kort beskrevet noe om det tematiske forholdet mellom HELVYRD og MECVET prosjektene. HELVYRD-studien med forskning på vurderernes helhetlige vurderinger har sitt utspring fra deler av MECVET-prosjektet. Videre er designet videreutviklet ved å inkludere to brukte vurderingssystemer fra fag- og yrkesopplæring (Figur 16).



Figur 16. Sammenhengen mellom de utvalgte vurderingssystemene

I utviklingen og utprøvingen av testverktøyet i MECVET-prosjektet kom det fram at spørsmål knyttet til vurdererernes vurderinger er interessante å følge opp som en egen aktivitet. Temaet formulert som «community of raters» var også blinket ut som viktig i prosjektsøknaden til MECVET.

##### 4.2.1 MECVET-designet og metodene

MECVET-prosjektet følger et metodisk opplegg som er vanlig innen utvikling av kompetansetester. Dels er det begreper og teknikker hentet fra psykometri og dels prinsipper for design av måleinstrumenter.

MECVET-prosjektet skal utvikle et instrument for å kunne vurdere og måle yrkeskompetanse med utgangspunkt i KOMET-modellen. Den er som nevnt teoribasert, og vil være prøvd ut på ulike yrkesgrupper eller fag uten store endringer i dimensjoner og kriterier. Utviklingen av måleinstrumentet involverer eksperter fra de respektive fagene, og foregår gjennom flere sykluser der kompetanseprofil og testoppgaver valideres med hensyn til innhold, og i noen grad format. Det inngår også en pilotprøving på et mindre utvalg elever og lærlinger i de

aktuelle fagene. Slike runder med instrumentutvikling gjennom iterasjoner er kjent fra pedagogisk designforskning uten at jeg vil bruke denne betegnelsen på mitt prosjekt.

I designet til MECVET er det gjort rede for syv ulike faser prosjektet skal igjennom. (1) Etablering av oppdatert teoretisk og metodisk plattform samt å informere/forankre konseptet inn i ulike instanser. (2) Validering av kjerneoppgaver i fire yrker (elektrikere, industrimekanikere, helsefagarbeidere og frisører (sistnevnte er utelatt)). (3) Utvikling av testoppgaver. (4) Kvalitative studier av vurderingsbaner og vurderernes bedømmelse gjennom å ha høyttenkning underveis og «stimulated recalls»-sesjoner. (5) Utvikling av en undersøkelse av kontekstuelle faktorer. (6) En pilotering av måleinstrumentet (testoppgaver og spørreskjema) og en bredere undersøkelse hos yrkesfagelever/lærlinger på Vg2 og første og andre års lærlinger. (7) Revidering av prosjektet. Figur 17 viser forskningsdesignet, der fasene oppfattes som fram- og tilbakeprosesser, og ikke et lineært forløp.



Figur 17. Forskningsdesignet MECVET (hentet fra MECVET-søknad til Norges Forskningsråd (NFR))

Punkt 4 har fokus på kvalitative studier av vurderernes vurderinger. Dette vil være utgangspunktet for utvikling av HELVYRD-prosjekt.

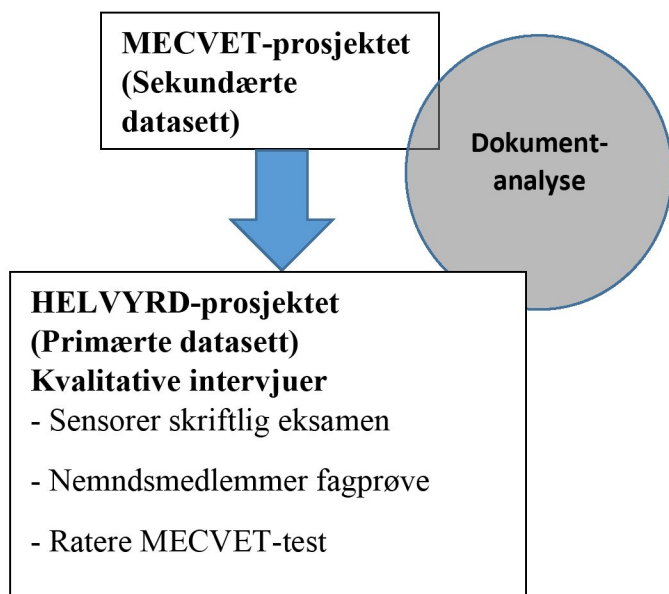
MECVET vil basere seg på både kvalitative og kvantitative forskningstradisjoner i utviklingen av verktøyet for kompetansemåling. I flere workshops skulle eksperter i de ulike fagene komme fram til en yrkesprofil, sentrale arbeidsoppgaver og læringsutfordringer i yrket. Her foregikk datainnsamlingen blant annet med støtte i video og lydopptak. Dessuten ble noen av møtene gjennomført på nett (Adobe Connect og Skype), og tatt opp på bånd. Fagekspertgrupper (6-8 stk) bestod av personer fra industrien, prøvenemnder, opplæringskontorer og lærere fra videregående skolen. Målet har vært å få grupper sammensatt fra ulike deler av landet bestående av dedikerte eksperter i sitt eget yrke. Disse er rekruttert av forskerne i MECVET-prosjektet ut ifra de nevnte kriteriene. Ekspertgruppene har blant annet i workshopene vært med på å definere og beskrive sentrale *kjerneoppgaver* i yrket som kjennetegner faget. Kjerneoppgavene skal ifølge KOMET-modellen være det empiriske grunnlaget (innholdet) for å utvikle *testoppgavene*. Testoppgaveutviklingen har gått over tid, og har vært en prosess, hvor man hele tiden justerer og evaluerer oppgavens innhold og relevans.

#### 4.2.2 HELVYRD-designet

Til forskjell fra MECVET, som er bygd etter en teoridrevet logikk der yrkeskompetanse operasjonaliseres i et sett av dimensjoner og kriterier for forhåndsbestemt koding, vil HELVYRD ha en mer eksplorerende karakter (se Halvorsen 2008 side 135). I hovedsak har mitt vitenskapsteoretiske ståsted i denne sammenheng vært knyttet til en kvalitativ tilnærming der fortolkning av resultater, utsagn fra intervjuer og dokumentdata er vesentlig, hvor prinsippene om en hermeneutisk fortolkning (Kvale & Brinkmann, 2015, s. 237) er en del av rammen.

Figur 18 illustrerer designet til HELVYRD. Dataene er i all hovedsak innhentet ved hjelp av personlige intervjuer med en eller to informanter (samtidig). Empirien fra MECVET vil fungere som sekundærdata for mitt prosjekt ved å ramme inn mine resultater i et større materiale, samtidig som de kvantitative dataene fra MECVET inngår som data for kvalitative analyser av mine data, for eksempel om det mine informanter sier de legger vekt på stemmer med vurderingene eller skårene i MECVET-materialet. Med flere vurderingssystemer i HELVYRD vil et slikt komparativt design begrunnes med at det kan gi en bredde som kan belyse ulike sider av vurderernes helhetlige vurderinger. En mindre del av datainnsamlingen består av dokumentanalyse knyttet hovedsakelig til oppgavetekster og tilhørende dokumenter.





Figur 18. Illustrering av hovedundersøkelsen og forholdet mellom det sekundære datasettet og primære datasettet. Det er selvsagt en glidende overgang mellom disse typene siden jeg har deltatt i innsamlingen av de sekundære dataene. Slik sett kunne forholdet og designet mitt være et eksempel på mixed methods der kvalitative og kvantitative data og analyse metoder integreres (Creswell, 2014, s. 14).

Valg av metoder har vært inspirert av tidligere forskning. Jeg har vist til forskere som brukte både «stimulated recall» og «think-aloud» i intervjuene. Det som nok skiller seg ut ved HELVYRD, er gjennomføringen av en relativt kort seanse med vurdererne rett etter at vurderingen er gjennomført, med ett oppfølgingsintervju omtrent en uke i etterkant. I tillegg valget av flere vurderingssystemer i samme studie brukt på lærlinger som er relativt like i sitt lærlingeforløp (slutten av Vg3).

#### 4.2.3 Hva svarer HELVYRD på kontra MECVET?

I utviklingen og gjennomføringen av datainnsamlingen for HELVYRD-prosjektet har utgangspunktet vært å studere og gå igjennom allerede eksisterende datamateriale som er innsamlet gjennom MECVET-prosjektet.

Som tidligere nevnt har MECVET-prosjektet samlet inn kvantitative og kvalitative data som utgangspunkt for å formulere spørsmål om testvurderernes vurderinger. Materiale av interesse er først og fremst video- og lydopptak av workshops med diskusjoner i ekspertgruppene omkring vurdering av besvarelser til elever og lærlinger, opptak fra diskusjoner i raternes (to og to) bedømmelser av besvarelser, samt noen få opptak av enkelt personer («think-aloud» - metoden). Målet i denne delen av analysen var å få et første innblikk i eksisterende datamateriale med de diskusjoner og refleksjoner som har pågått rundt oppgaver og vurderinger av besvarelser. Dette arbeidet ga en del empiri som kunne belyse mine

forskningsspørsmål, samtidig som det ga mer informasjon i den videre utviklingen av forskningsdesignet til HELVYRD. Dette kan ses på i et metodetriangulerings-perspektiv (Concepts of triangulation) hvor ulike metoder som intervju og gruppediskusjoner brukes for å få et bedre tolkningsgrunnlag (Flick, 2007, s. 6). Sånn sett vil dette kunne gi fyldigere materiale som styrker mitt design. Materialet fra MECVET fanger opp deler av mine spørsmål knyttet til spørsmålet om hva de legger vekt på i sine vurderinger, og hvordan det er å være en rater i dette systemet. Men materialet innehar ikke mulighet for å koble det raterne diskuterte i vurderingen opp imot konkrete besvarelser. Videre var opptakene kun knyttet til de samtalene og diskusjonene som ble tatt opp under workshopene. Det ble i mindre grad stilt spørsmål rundt deres vurderinger. I HELVYRD er det gjennomført direkte kvalitative intervjuer med vurdererne hvor «stimulated recall» og «think-aloud» ble benyttet. Hensikten er å komme dypere inn i vurdererens tenkning rundt deres vurderingsarbeid. Dette var vesentlig for å få innblikk i hva de la vekt på i de ulike besvarelsene, og hvordan de foretok vurderingene, for eksempel om de brukte formulerte kriterier, eksempler på gode og dårlige oppgaver.

Jeg skal videre gjøre rede for hvordan intervjuene ble utviklet og gjennomført i HELVYRD.

#### 4.3 Generelle metodiske prinsipper for intervjuene i HELVYRD-prosjektet

Å utføre intervjuene innebærer at det er noen overordnede generelle prinsipper og retningslinjer som er lagt til grunn. Jeg vil først redegjøre for dette før jeg går nærmere inn på utvalget av informanter, utvikling av intervjuguider og gjennomføring av intervjuer i hvert enkelt vurderingssystem.

Det å gjennomføre et forskningsintervju er en interpersonlig situasjon. To parter har en samtale med felles interesse om et bestemt emne (Kvale & Brinkmann, 2015, s. 156; Patton, 2015, s. 36). I intervjuet skapes det kunnskap i skjæringspunktet mellom intervjuerens og den intervjuedes synspunkter (Kvale & Brinkmann, 2015, s. 83). For at intervjuet skal bli en seanse hvor dette kan skje, er det viktig med en god iscenesettelse (Kvale og Brinkmann, 2015, s. 160) (slik at den intervjuede føler seg komfortabel og trygg i situasjonen).

I den praktiske gjennomføringen av intervjuene i HELVYRD ble følgende lagt vekt på:

- Sikre at den enkelte vurderer har fått all informasjon og forstått hva prosjektet går ut på, samt sikre at vedkommende har signert informert samtykke (informasjonsskriv).
- Tidsrammen på hele intervjuet pr. vurderer er omtrent en time. Dette vil variere noe ut fra

den enkelte vurderer.

- Alle intervjuene ble tatt opp på lydbånd og transkribert av meg.
- Intervjuene tar utgangspunkt i de konkrete oppgavetekstene og besvarelsene fra lærlingene.
- Så langt det er mulig skal det etterstrebtes å få besvarelser med forskjellige karakterer som dekker gjeldende «karakter-skala». Skalaen er ulik for de tre vurderingssystemene. Hensikten er å få fram vurderernes syn ved de ulike vurderingsnivåene.
- Hovedstrukturen bygger på å intervjuer aktuelle informanter i en kombinasjon av en og en og to og to. Men vurderingene gjennomføres på ulike måter i de respektive vurderingssystemene, og krever noe ulik tilpasning når det gjelder tid og tilgjengelighet.
- I slutten av intervjuet er det viktig med en debriefing for å gå igjennom hva man har diskutert.

Intervjuene er lagt opp slik at de intervjuede blir bedt om å forholde seg til konkrete besvarelser fra lærlinger. Dermed får man skapt situasjoner hvor det er lettere å gjenkalle og å ha oppmerksomheten mot viktige detaljer i vurderingene som er blitt gjort. Denne formen for intervjuer minner om «stimulated recall» (Calderhead, 1981, s. 54; Lyle, 2003, s. 863). Lyle (2003, s. 873) påpeker at en «stimulated recall»-seanse bør skje i en kontekst som er nærmest mulig den angitte situasjonen. Lyle (2003, s. 872) peker på «The main concern is the extent to which the subjects have reordered their thoughts before or during the recall process» Lyle (2003, s. 872). I HELVYRD-studien er det lagt opp til to runder med intervjuer med noe tid i mellom disse, og Lyle (2003, s. 872) peker på at dette kan gi en mulighet for å se om tankene på det opprinnelige i intervjurunde en har endret seg (reordered) ved runde to. Lyle (2003, s. 873) gjør et poeng av at det må trekkes distinksjoner mellom ulike måter å bruke «stimulated recall» på. En metode er å få fram de tankene som var der da en avgjørelse skulle tas, og en annen hvor «stimulated recall» brukes for å oppmuntre til refleksjon. I min studie vil begge situasjoner være ønskelig for å få den umiddelbare tenkningen til de som vurderer samt en oppfølging i etterkant. Men det å validere graden av nøyaktighet på kognitiv «output» er svært vanskelig noe som må diskuteres som begrensninger i bruken av denne metoden. Derfor vil jeg kalle det for intervjuer med *stimulert oppmerksomhet* for å unngå vanskelige situasjoner om hvorvidt det som sies er en gjenkalling eller en vurdering som oppstår under intervjuet.

En mindre del av datainnsamlingen vil være knyttet til bruk av «tenke -høyt»-metoden. Patton (2015, s. 486) uttrykker det som «... aims to elicit the inner thoughts or cognitive processes that illuminate what's going on in a person's head during the performance of a task ...» (Patton, 2015, s. 486). Den kan benyttes fortløpende eller i et retrospektiv perspektiv (Sütő, 2012, s. 25), men Kuusela og Paul (2000, s. 397) uttaler at fortløpende «tenke-høyt» er mer

effektiv enn retrospektive metoder. I HELVYRD-studien skjer det fortløpende. «Tenke -høyt» er ikke passende for alle typer av datainnsamling og må tilpasses de ulike aktivitetene metoden er ment å kartlegge (Lyle, 2003, s. 862).

Greatorex og Sütő (2008, s. 319) har i sin forskning involvert vurderere av eksamener etter at de hadde deltatt i en studie ved hjelp av «tenke -høyt», og spurte dem om hvordan metoden påvirket deres tenkning om oppgaver. Greatorex og Sütő (2008, s. 329) har i sin begrensede studie oppsummert med at det var delte meninger i blant de 12 vurdererne. Noen mente at en «tenke -høyt» ikke hadde noen effekt på deres vurdering, mens andre igjen rapporterte at de merket en påvirkning, og stilte spørsmål ved bruk av denne metoden. Hva som opplevdes som lett og vanskelig å verbalisere var også delt, og det var ulike subjektive meninger om dette. De forskjellige svarene fra vurdererne kan bunne i ulike former for vurdering, og om det er forskjeller i deres egen selvbevissthet og innsikt. Videre reises det spørsmål om det er vanskeligere eller enklere å forklare de tankene som er enklest og mer automatisk enn de mer kompliserte tankene. Det denne studien konkluderer med, selv med et så lite utvalg, er at forskningsmetodisk er det begrensninger knyttet til å ha denne som en «stand-alone» (s. 330). I HELVYRD-studien vil metoden brukes i mindre grad, og det vil være sammen med andre metodiske tilnærminger.

I utformingen av kvalitative intervjuer viser Kvale og Brinkmann (2015, s. 137) til syv stadier for gjennomføring: tematisering, planlegging, intervjuing, transkribering, analysering, verifisering og rapportering. De innledende fasene har jeg vært innom tidligere. Men her skal jeg se nærmere på selve gjennomføringen av intervjuet.

#### 4.3.1 Utvalg av informanter

Utvalget av informanter vist i figur 19 fordelt på vurderingssystemer og nemnder, sensorgrupper og ratergrupper.

<b>Vurderingssystemer (Elektrikerfaget)</b>	<b>Antall informanter</b>	<b>Kommentarer</b>
Nemndsmedlemmer fagprøve	5 stk	Fordelt på tre nemnder
Sensorer skriftlig eksamen	4 stk	Fordelt på to sensorgrupper
Ratere MECVET-tester	4 stk	Fordelt på to ratergrupper

Figur 19. Oversikt på informanter fra de ulike vurderingssystemene

Utvelgelsen av vurdererne knyttet til HELVYRD-prosjektet er først og fremst basert på en idé om hvilke forskjeller som er viktige å få med i utvalget, slik som faglig bakgrunn og deltakelse i de ulike vurderingssystemene (tilsvarende kvoteutvelgelse innen kvantitativ metode). En del kriterier som erfaring i vurderingsarbeid og kjønn fikk ingen avgjørende betydning, og her var det også nødvendig å ta hensyn til tilgjengelighet.

De fem utvalgte nemndsmedlemmer er tilknyttet samme prøvestasjon. Fagprøver i elektrikerfaget avlegges i hele landet både i bedrifter og på prøvestasjoner. Jeg valgte en ekstra nemnd hvor det ene nemndsmedlemmet deltok i to av nemndene. På det tidspunktet (januar og februar 2015) hvor intervjuene ble gjennomført, var to av nemndsmedlemmene knyttet til næringslivet, mens to andre var knyttet til både skole og næringsliv, hvor den ene av disse deltok også som rater i MECVET-prosjektet, og hadde tilhørighet til yrkesfaglærerutdanningen på Høgskolen i Oslo og Akershus. Den siste hadde nettopp begynt å jobbe i skolen. Fire av dem har vært med i prøvenemnd i en rekke år og har lang erfaring, mens det femte medlemmet hadde vært med noen få år.

Når det gjelder skriftlige eksamener, er det en sentral nemnd som lager en prøve for hele landet. Det blir oppnevnt sensorer i de fylker hvor det er praksiskandidater/lærlinger som gjennomfører eksamen. Utvalget av sensorer kommer fra samme fylke, og de samarbeider om å sensurere besvarelsene. Jeg vil presisere at det kan være store variasjoner fra fylke til fylke hvordan sensureringen gjennomføres. På et tidspunkt hvor intervjuene ble gjennomført (juni 2015), jobbet tre av sensorene som lærere i skolen, mens den siste jobbet i næringslivet. Alle har vært med som sensorer i en rekke år.

Når det gjelder ratere tilknyttet MECVET-testene, er utvalget basert på tilgjengelighet blant de få som er knyttet til elektrikerfaget og MECVET-prosjektet. På det tidspunktet dette ble gjennomført (Februar 2015), var det, i tillegg til nevnte informant fra fagprøven, en av

deltakerne som jobbet som lærer i skolen. En jobbet på opplæringskontor for Elektrofag, og har også vært tilknyttet yrkesfag Elektro på yrkesfaglærerutdanningen ved Høgskolen i Oslo og Akershus. Den siste jobbet i næringslivet, og var utøvende elektriker.

Både antallet og kategorien informanter er bestemt ut fra en avveining mellom strategisk utvelgelse (forskningsspørsmål, hovedtyper m.m) og praktisk hensyn. Det strategiske (purposeful sampling) (Patton, 2015, s. 311) ligger i å få representert de enhetene (fenomenene) man vil studere. I mitt tilfelle er denne studien delt i to hvor det er et sett med vurderere knyttet til hvert enkelt vurderingssystem (som enkelt-case), samtidig som vurderingssystemene sammenlignes med hverandre (multi-case). Innenfor hvert enkelt case (vurderingssystem) har jeg forsøkt å bruke kriterier knyttet til faglig bakgrunn, det vil si kriterieutvelgelse der hensynet har vært å få variasjon. Det samme gjelder utvelgelse av test-, prøve og eksamensbesvarelser (se nedenfor). I praksis har jeg måttet ta hensyn til tilgjengelighet, men med utgangspunkt i kriteriene vet jeg at viktige dimensjoner er representert i dataene. Når det gjelder representativitet, ble kombinasjonen totalt sett god. Det var næringslivsfolk med lite og mye erfaring rundt vurdering. Det var lærere fra videregående skole og fra opplæringskontor. Begge kjønn ble representert, dog med stor andel av det mannlige (noe som ikke er uvanlig i elektriker yrket). Miksen av representanter er mangfoldig, og deltakerne kommer i fra sentrale deler av opplæringssystemet (bedrift, opplæringskontor og videregående skole). Denne fordelingen gjaldt ikke innad i alle tre vurderingssystemene, men totalt for alle tre. For utvalget i sin helhet vil jeg si det er typisk for det de skal utføre av vurderingsarbeid (Halvorsen, 2008, s. 164). Halvorsen viser til at man først gjør en analytisk generalisering ved å velge ut prøvesystemer og individer som antas å gi et rikt bilde av det man skal studere. Representativitet i utvalget betyr at informantene man velger er typiske for det man skal finne ut av. Det er rimelig å anta at mitt utvalg er forholdsvis typisk for hver av kategoriene (personer, oppgaver, vurderingssystemer), og at denne likheten muliggjør generalisering (transferability) (Denzin & Lincoln, 2018, s. 20), samtidig som jeg vil være oppmerksom på særegenheter i hvert enkelt tilfelle. Utvalget av dokumenter (tester og prøver) vil også kunne vurderes som mer eller mindre typiske for eksempel langs en karakterskala. Også datasettet for MECVET gjør det mulig å si noe om hvorvidt utvalget av elektrikerbesvarelser er representative for et langt større utvalg fra samme fag (N=150). Bruk av «mixed methods» vil kunne redusere kravet til antall informanter for kvalitativ intervjuing (Charmaz, 2014, s. 105). Det må også nevnes at studien er eksplorerende fordi det er gjort meget få studier av vurderernes skjønnsutøvelse innen fag-

og yrkesopplæring. Derfor vet man ikke på forhånd hvilke faktorer som er viktige å få representert i utvalget. I slike tilfeller er det viktig å være tydelig (transparent) på hvordan man har kommet fram til det aktuelle utvalget, slik at eventuelle oppfølgingsstudier kan bygge videre på min studie.

Proessen med å rekruttere deltakere og å få organisert og gjennomført intervjuene har til dels vært krevende. Utfordringen i de ulike vurderingssystemene har vært tilgang og timing med informantene for å få gjennomført intervjuene. Å finne informanter er en øvelse som krever tid og tålmodighet, hvor dette ofte skal godkjennes på andre nivåer i, som for eksempel fylker. Ambisjonen var å få noe geografisk spredning på de ulike vurderingsgruppene (de norske), men tid og ressurser medførte at det ble avgrenset til nærområdene.

Innstillingen til å delta hos informantene har vært udelt positiv. Mitt utvalg av vurderere knyttet til de tre vurderingssystemene er begrenset i antall innad i systemene, og man skal være forsiktig med å generalisere noe ut ifra den empirien som er samlet inn. Men ut fra det eksisterende utvalget vil innsamlet data blant disse informantene kunne gi et godt bidrag i å svare på forskningsspørsmålene. Begrepsvaliditeten har jeg forsøkt å styrke ved å presisere de teoretiske begrepene slik at de kan gjenfinnes i mitt materiale. Innholdsvaliditeten er vurdert i samtale med medforskere i MECVET og informanter, mens reliabiliteten er understøttet gjennom systematikk i kodingsarbeid, gjentatte fortolkninger og samarbeid med kolleger. Man kan innvende til utvalget at det kan være store variasjoner i nemnder og blant sensorer over det ganske land som ikke fanges opp av mitt design. Skulle utvalget vært utvidet med flere, ville det være naturlig å involvere informanter fra flere nemnder og sensorgrupper i andre deler av landet, noe som kunne gitt en ekstra dimensjon på ulikheter og likheter.

#### 4.3.2 Utvikling av intervjuguider og intervju spørsmål

Ved gjennomføringen av intervjuene basert på *stimulert oppmerksomhet* vil spørsmålene være semi-strukturerte og utformet som temaer. Halvorsen (2008, s. 137) uttaler at ustrukturerte intervjuer er interessant for å finne ut av fenomener som allerede har skjedd, hvor forskeren ikke har observert selv. Intervjuene er utformet med utgangspunkt i en problemstillingen, dessuten en liste over de temaer som skal tas opp. Et intervju spørsmål kan ha en tematisk og en dynamisk dimensjon. Den tematiske (hva) handler om produksjon av kunnskap, og den dynamiske (hvordan) om den interpersonlige relasjonen i intervjuet (Kvale og Brinkmann, 2015, s. 163). Gjennom å ikke ha et for stivt opplegg vil man kunne få belyst problemstillingen fra ulike synsvinkler som kan gi en fyldigere og mer helhetlig bilde.

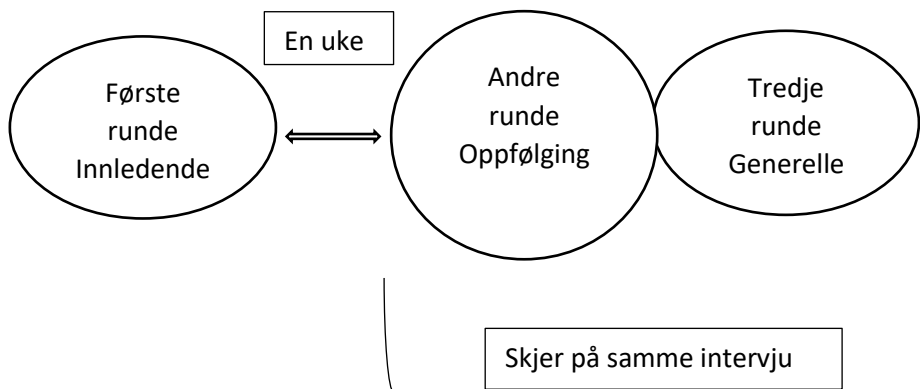
Halvorsen (2008, s. 138) sier om dybdeintervjuer at disse er uformelle og følger samtalsgang, men ofte har intervjuer en liste med ulike temaer som skal diskuteres. Fordelen er at den intervjuede ikke tvinges inn i en bestemt måte å tenke på. Man kan uttrykke og utdype sine meninger mye friere. Min variant er en blanding av den ustrukturerte typen og dybdeintervjuet, altså mer semi-strukturert.

Kvale og Brinkmann (2015, s. 165) er innom flere typer av intervju spørsmål hvor jeg med mine spørsmål vil inneha flere av disse i form av blant annet *introduksjon, oppfølging, inngående, spesifiserende, strukturerende, og fortolkende*. Oppfølgende spørsmål vil bli sentralt når man skal kartlegge noe som kan være en form for «taus kunnskap» hos vurdererne. Målet med intervju spørsmålene er å komme dypere inn i hva og hvordan vurdererne gjør sine vurderinger. Hensikten er å få fram mest mulig informasjon i en intervjusituasjon hvor vurdererne er i en kontekst som er nærmest mulig den aktuelle vurderingssituasjonen. Sett opp mot «hvorfor»-spørsmålet i utviklingen av intervju spørsmålene har jeg vært bevisst på i minst mulig grad å benytte dette ordet direkte. Patton (2015, s. 455) tar opp ulike sider ved «hvorfor»-ordet som kan være med på å mistenkeliggjøre informantens svar. Patton uttaler «When used as a probe, «why» questions can imply that a person's response was somehow inappropriate» (Patton, 2015, s. 455). Bruken av ordet kan utløse tvil til det respondenten uttrykte. Skal man komme noe dypere, er det viktig å bruke andre formuleringer som kan få fram mer informasjon om temaet. Min tanke er å la informanten være i en situasjon hvor de mer umiddelbare tankene kommer, og igjennom den prosessen vil de kunne gi noen innspill til forståelse av «hvorfor»-spørsmålet omkring deres vurderinger.

Intervjuguidene (vedlegg 1) for de tre vurderingssystemene er inndelt i tre runder (figur 20). Den første runden vil være intervju spørsmålene, som danner grunnlaget for å få fram det umiddelbare hos vurdererne rett etter at vurderingen er gjennomført. I den andre runden (ca. en uke etter) gjennomføres oppfølgings spørsmål tilpasset den enkelte vurderer basert på datamateriale fra første runde, og den gå noe mer i dybden. Vurdererne fikk ikke tilgang til utskrift fra første runde fordi jeg ønsket å utforske videre med tenkningen om stimulated recall. Men innledningsvis i runde to ble det gjort en repetering og kortfattet oppsummering av runde én. Oppfølgings spørsmålene vil variere i form ut fra hva som kom fram i første runde. I den tredje runden, umiddelbart etter den andre runde, vil intervjuet ta for seg noen generelle spørsmål rundt vurdering. Disse spørsmålene er utviklet basert på en del av min gjennomgang av forskningen på feltet og datamaterialet fra MECVET-prosjektet. Dette



innebærer en viss struktur på intervjuet hvor alle deltakerne svarer på mange av de samme spørsmålene. I mitt forskningsdesign har jeg også lagt vekt på at denne studien også er en kombinasjon av å ha personsentrerte tilnærminger som vil gi perspektiver som er mer helhetlige og sikre at temaene som blir valgt ut ikke blir for oppsplittet. Dette gir også en mulighet for å se de utvalgte temaene i en sammenheng for hver enkelt person (Thagaard, 2013, s. 181).



Figur 20. Gangen i gjennomføringen av intervjuene

Figur 21 viser sammenhengen mellom forskningsspørsmålene og intervju spørsmålene.

	Tema	Datatilfang/intervju spørsmål	Kommentarer
Forskningsspørsmål 1	Helhetlig vurdering Yrkesdidaktikk	Dokumentanalyse Generelle spørsmål (noen) Litteratur	Sekundære data vil inngå som en del i alle faser
Forskningsspørsmål 2	Gjennomføring av vurdering Vurderernes kultur	Dokumentanalyse Første runde innledende spørsmål Andre runde oppfølgingsspørsmål Generelle spørsmål (noen)	
Forskningsspørsmål 3	Påvirkning av ytre faktorer Vurderingssystemer og oppgaver	Dokumentanalyse Første runde innledende spørsmål Andre runde oppfølgingsspørsmål Generelle spørsmål (noen)	

Figur 21. Sammenhengen mellom forskningsspørsmålene og metoder for innhenting av data.

Intervjuguidene blir noe tilpasset når det gjelder spørsmålene i hvert enkelt vurderingssystem. Organiseringen og gjennomføringen av intervjuene varierer noe i form ut fra hvilke praktiske

muligheter som er tilstede. Dette fikk noe betydning for spørsmålsutformingene i intervjuguidene, og jeg måtte ha en pragmatisk tilnærming til konteksten intervjuene ble gjennomført i.

Hovedprosessen med gjennomføring av intervjuene vil være likt, men noen deler vil skille seg ut. Gjennomføringen av intervjuene for hvert vurderingssystem vil bli redegjort for. I den prosessen kan det oppstå noen gjentakelser. Dette medfører at tilfanget av data vil være samlet inn på noe ulike måter i henhold til de forskjellige informantene relatert til de ulike systemene. Jeg har stort sett brukt de samme prosedyrene for datainnsamling for hvert av vurderingssystemene slik at jeg kan gjøre «fler-case» sammenligninger.

Pilotering av intervjuguiden (Kvale og Brinkmann, 2015, s. 151; Halvorsen, 2008, s. 147) vil jeg si i mitt tilfelle har skjedd gjennom flere faser i utviklingen av temaer og intervju spørsmål. Jeg har støttet meg til tidligere forskning som har gitt nyttige innspill. For det andre, som nevnt tidligere, har datamaterialet fra MECVET-prosjektet vært et viktig bidrag til utformingen av relevante spørsmål. For det tredje har det vært flere runder med veiledninger og diskusjoner rundt de aktuelle intervju spørsmålene. For det fjerde var jeg til stede i forkant av selve intervjudagen for å observere og snakke med folk om gjennomføringen av fagprøven. Dette ga viktige innspill og tips til intervjuguiden. For det femte er en av grunnene til at jeg valgte tre nemnder når det gjelder fagprøven at intervjuene av den første nemnda ville kunne fungere som en test (pilotering), og skulle noe skje, hadde jeg mulighet for å korrigere. Det viste seg å ikke være nødvendig. For det sjette er intervjuguiden bygd opp slik at det som eventuelt blir glemt, eller skulle vært diskutert mer i første runde, kan bli fanget opp og tatt opp i runde to og eventuelt runde tre. Til slutt ga intervjuene i første vurderingssystem (fagprøven) erfaringer som ville komme til gode ved gjennomføringen av intervjuer i de to resterende vurderingssystemene.

#### 4.4 Gjennomføring av intervjuer

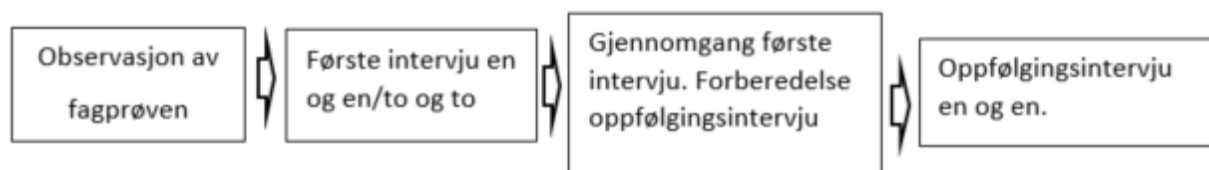
I presentasjonen av intervjuene knyttet til de respektive vurderingssystemene vil dette skje i den rekkefølgen intervjuene ble gjennomført: Fagprøven og nemndsmedlemmer (januar og februar 2015), MECVET-test og ratere (februar 2015) og eksamen og sensorer (juni 2015). Til slutt kommer en oppsummering av de erfaringene som ble gjort.

#### 4.4.1 Intervju av prøvenemndsmedlemmer til fagprøven

I forkant (høsten 2014) av datainnsamling ble jeg invitert til et møte med alle prøvenemndene. Der hadde jeg en 20 min lang presentasjon av HELVYRD-prosjektet, og hadde en diskusjon om interesse for deltakelse. Resultatet ble at jeg skulle møte opp til prøvestasjonen i (januar og februar 2015) hvor mange lærlinger skulle ta fagprøven. Ved møte i januar var jeg i kontakt med flere nemnder. Ut fra praktiske hensyn og interesse for å delta, ble tre nemnder valgt. Det ble avtalt uker, to av nemndene var i samme uke og den tredje to uker senere. I forkant av dette ble jeg med som observatør en uke for å bli mer kjent med gjennomføringen av fagprøven. Enkelte deler av fagprøven ble observert, men av naturlige grunner var det ikke mulig å delta på alle faser. Jeg kommuniserte ikke med lærlingene. Uformelle samtaler oppstod med enkelte prøvenemndsmedlemmer, som ga informasjon om fagprøvens ulike sider. Dette ga større innsikt i hvordan fagprøven ble utført, og ga meg et bedre grunnlag i prosessen med å utvikle spørsmål for min intervjuguide.

I hver nemnd er det to nemndsmedlemmer, og en av personene var med i to nemnder, slik at jeg samlet intervjuet fem personer. Hver nemnd hadde per uke fire lærlinger de skulle vurdere. Totalt ble det gjennomført fagprøve for 11 lærlinger i de tre nemndene jeg fokuserte på. Det var en lærling som ikke gjennomførte. Det ble en jevn spredning på de karakterer som ble gitt, hvor alle tre nivåer var representert (ikke bestått – bestått – bestått meget godt). All dokumentasjon knyttet til fagprøven ble samlet inn og skulle danne grunnlaget for intervjuet og gjenerindring av vurderingsprosessen.

Hovedgangen i gjennomføringen av intervjuene ble fulgt slik det ble presentert i figur 20. Figur 22 viser mer konkret gjennomføringen for fagprøven.



Figur 22. Gjennomføring av intervjuene med prøvenemndsmedlemmene ved fagprøven.

Det første intervjuet ble gjennomført rett etter at vurderingsarbeidet var ferdig. De to nemndsmedlemmene ble først intervjuet en og en. De hadde vurdert fire lærlinger, og dokumentasjonen lå på bordet. Jeg ønsket å få den umiddelbare tenkingen til nemndsmedlemmet om den vurderingen som var gjort på de fire lærlingene. Det startet med et innledende spørsmål hvor nemndsmedlemmet fikk prate. Stort sett hadde de mye på hjertet,

og mye ble sagt på relativt kort tid. Tiden brukt i denne fasen var ca.15 min. pluss/minus pr. nemndsmedlem på de fire lærlingene. Etter de individuelle intervjuene ble de to nemndsmedlemmene samlet til et felles intervju. Noen av de samme spørsmålene ble stilt om igjen. Dette ga mulighet for noe mer utdyping samt å avdekke enigheter og uenigheter i diskusjonene mellom nemndsmedlemmene. Denne delen av intervjuet kunne vare opp til 20 min. Det ble gjort likt med alle tre prøvenemndene.

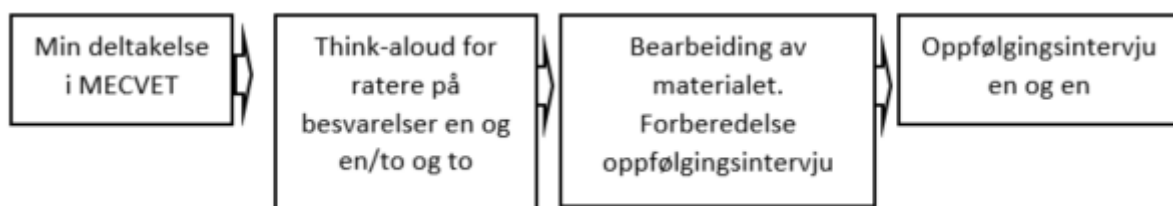
Materialet produsert etter den første runden ble gjennomgått og jeg laget oppfølgingsspørsmål til det enkelte nemndsmedlemmene. Tidsavstanden mellom første runde og andre runde var begrenset til maks en uke. I tilfellet med nemndsmedlemmer var det viktig å få gjort dette før de startet opp med en ny gruppe lærlinger. Dette kunne medføre at de glemte det de hadde vært igjennom med disse lærlingene.

Den andre delen av intervjuet ble gjennomført en og en. En time var satt av og intervjuene kunne vare både kortere og lengre. Intervjuene ble gjennomført basert på oppfølgingsspørsmål til det enkelte nemndsmedlem. Til slutt ble det stilt mer generelle spørsmål om deres vurderingspraksis.

#### 4.4.2 Intervju av ratere på MECVET besvarelser

Ved gjennomføringen av intervjuene av raterne fikk de tilsendt informasjon på mail om hvordan de skulle gå fram for å gjennomføre ratingen. Åtte besvarelser ble valgt ut fra 2. års lærlinger med tilhørende ratingskjemaer fordelt på alle de fire oppgavene. Utvalget er basert på det som var tilgjengelig av besvarelser på det aktuelle tidspunktet. Det kom flere besvarelser høsten 2015, men dette ble for sent i progresjonen til HELVYRD-studien.

Det var ønskelig å finne besvarelser og ratinger som dekker spennet i de ulike nivåene: *ikke oppfylt – dårlig oppfylt – nesten oppfylt – oppfylt*. I dette tilfellet var det ikke spredning på alle nivåene, kun et par unntak, og de fleste lå rundt de to laveste nivåene. Dette kjennetegnet de fleste besvarelsene fra de lærlingene som deltok. Figuren 23 viser gangen i gjennomføringen av intervjuene med raterne.



Figur 23. Gjennomføringen av intervjuene med ratere ved MECVET-testen

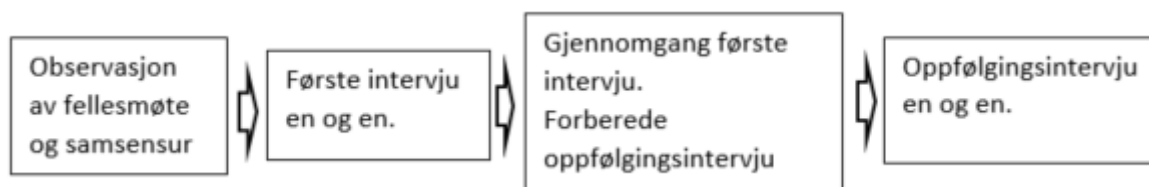
De fire raterne ble delt inn i grupper på to og to. Hver ratergruppe fikk tilsendt fire besvarelser, hver fordelt på to ulike oppgaver. Det var ikke praktisk mulig å samle disse raterne til umiddelbare intervjuer etter vurderingene, så det ble organisert for gjennomføring «think-aloud». I den første fasen satt raterne hver for seg hvor de pratet inn på lydopptaker når de gjorde sine vurderinger. Neste fase i prosessen møttes de to og to for gjennomføring av omforente vurderinger. Her skulle de også gjøre innspilling av diskusjonene. Det ble ikke gitt noen detaljerte føringer på hvordan dette skulle gjøres utover det at de ble bedt om å ta lydbåndopptak under sin ratingprosess. Alt materialet fra første runde ble overlevert til meg som lydfiler og ratingskjemaer (individuelle og omforente). I gjennomgangen av materialet lyttet jeg til lydfilene og fulgte vurderingsarbeidet samtidig som jeg satt med ratingskjemaene foran meg. Dette ble gjort både for den individuelle og omforente vurderingen, og dannet grunnlaget for å utvikle spørsmål til oppfølgingsintervjuet.

Ved oppfølgingsintervjuene møtte jeg den enkelte rater ut i fra hvor det var mest praktisk og treffes og hvor rammene rundt var tilfredsstillende. Intervjuene varte i snitt ca. en time pluss/minus. Grunnet prosessen i forkant med «think-aloud» valgte jeg i denne runden innledningsvis å stille noen innledende spørsmål for å fange opp litt av det som ble sagt på lydbånd. Deretter gikk jeg mer inn på de konkrete oppfølgingsspørsmålene og tilslutt de mer generelle.

#### 4.4.3 Intervju av sensorer på sentralgitt skriftlig eksamen

I forkant av intervjuene med sensorene hadde jeg kontakt med koordinator for denne gruppen pr. mail, hvor ulike forhold som organisering og gjennomføring ble avklart. Selve tidspunktet for gjennomføringen av intervjuene av de fire sensorene startet med at jeg ble invitert til et fellesmøte (se figur 24). De hadde en felles gjennomgang og diskusjon omkring selve oppgaven. Jeg inntok en observatørrolle i denne delen og gjorde notater. I forkant av møtet hadde sensorene hver for seg gjort sine vurderinger av besvarelsene. Møtet varte omlag en time og kan ses på som starten på samsensurering av besvarelsene. Etter dette gikk de sammen to og to, hvor det ble en kortere økt med diskusjoner om de ulike besvarelsene og karakterene. Denne seansen varte ikke lenge og det så ut til at de ble raskt enige.

Etter runden med sensurering ble hver sensorgruppe bedt om å ta med besvarelser som dekket de ulike karakterene så langt det lot seg gjøre. Det endte med tre besvarelser fra hver gruppe som til sammen dekket hele karakterskalaen. Grunnlagsmaterialet for intervjuene var selve oppgaven og besvarelsene.



Figur 24. Gjennomføringen av intervjuene med sensorene ved skriftlig eksamen

En faktor som spilte inn var knapphet på tid. I denne delen ble det kun organisert som ett en og en-intervju, og det ble ikke mulighet for intervju med sensorparene samlet. Det ble avtalt tidspunkter for oppfølgingsintervjuer som ble gjennomført omlag en uke etter. Jeg reiste til det stedet det passet for den enkelte. To stykker ble gjennomført på det stedet vi møttes første gang, og de to andre ble gjennomført på deres respektive arbeidsplasser. Intervjuene ble gjennomført som «normalt» slik det ble gjort i de andre vurderingssystemene. I snitt satt jeg en time, men det varierte basert på hvor mye intervjuobjektet snakket.

#### 4.4.4 Oppsummering og drøfting rundt gjennomføring av intervjuene

Figur 25 viser en oversikt på de ulike fasene i gjennomføringen av intervjuene i de tre ulike vurderingssystemene, som kan innebære likheter og ulikheter. I gjennomføringen er det forskjeller som bunner i at ulike hensyn må tas i intervjuene av informantene. Vil dette ha betydning for validiteten og reliabiliteten? Og kan dette få noe innvirkning på den empirien som samles inn?

<b>Fase</b>	<b>MECVET-tester Ratere</b>	<b>Fagprøver Prøvenemndsmedlemmer</b>	<b>Skriftlig eksamen Sensorer</b>
1	Deltatt på ratertrening i forkant i regi av MECVET	Observasjon og samtaler med nemnder på prøvestasjon i forkant	Observasjon av et felles sensormøte samt samsensur
2	Think-aloud en og en med opptak lydbånd	Intervju umiddelbart med en og en (stimulert oppmerksomhet)	Intervju umiddelbart med en og en (stimulert oppmerksomhet).

3	Think-aloud to og to med opptak lydbånd	Intervjuer to og to rett etter ovennevnte (stimulert oppmerksomhet)	Ikke mulighet med to og to i ettertid. Observerte i forkant og diskuterte det i en og en intervju
4	Bearbeiding av data og forberedelse oppfølgingsintervjuer	Bearbeiding av data og forberedelse oppfølgingsintervjuer	Bearbeiding av data og forberedelse oppfølgingsintervjuer
5	Oppfølgingsintervjuer (stimulert oppmerksomhet) med en og en	Oppfølgingsintervjuer (stimulert oppmerksomhet) med en og en	Oppfølgingsintervjuer (stimulert oppmerksomhet) med en og en

Figur 25. De ulike fasene i gjennomføring av intervjuene i de tre vurderingssystemene

Vurderingssystemene er ulike når det gjelder oppgaveutforming, gjennomføring og metode for vurdering. Dette medfører at intervjuguider må tilpasses, slik at de får betydning i det aktuelle vurderingssystemet. Men hovedrammen for intervjuguiden ligger som et utgangspunkt. I analysen må jeg ta høyde for disse forskjellene i opplegget.

Man skal ikke glemme betydningen av den spontaniteten det ligger i «think-aloud»-prosesser, hvor vurdererne umiddelbart taler inn sine tanker og forteller hva de fortløpende gjør. I et intervju med *stimulert oppmerksomhet* får man ikke fram den direkte spontaniteten, men samtidig er tiden kort mellom avslutningen av vurderingen og gjennomføringen av intervjuene. Jeg opplevde i de sistnevnte intervjuene at informantene fortsatt var langt inni tenkningen om hvorfor de gjorde de vurderingene de gjorde, og at de relativt spontant fortalte om hva de hadde lagt vekt på i sine bedømmelser. I tillegg kom fase tre, hvor vurdererne satt sammen og diskuterte besvarelsene, men hvor det i større grad kom inn refleksjoner. Sistnevnte mulighet hadde ikke sensorene ved eksamen på grunn av knapphet på tid, men seansene i forkant ga meg et innblikk i hva de diskuterte og hvordan de samsensurerte.

En vesentlig forskjell er knyttet til at ratere ved «think-aloud» var overlatt til seg selv. Empirien framstår annerledes her enn ved de umiddelbare intervjuene. Jeg hadde heller ikke

anledning til å stille noen spørsmål underveis. For prøvenemndsmedlemmene og sensorene ble det innledet med at de skulle fortelle hva de hadde lagt vekt på i sine vurderinger, og underveis kunne det stilles noen oppfølgingsspørsmål. Samtidig kunne dette bli fanget opp igjen i fase fire og fem med raterne, og oppfølgingsintervjuene ble noe annerledes enn for de to andre, siden starten ble mer lik starten for de umiddelbare intervjuene i fase 2.

I fase fire var det en gjennomgang av lydbåndopptak hvor det ble gjort notater underveis som dannet grunnlaget for å skreddersy spørsmål til den enkelte vurderer. Målet var å få fram mer data som kunne begrunne og belyse de valgene vurdererne gjorde. Lydbåndopptakene var forskjellige gjennom «think-aloud»-prosesser og *stimulert oppmerksomhet*-prosesser, men alle ga et innblikk i hvordan vurdererne tenkte. Det jeg merket var at empirien innsamlet via «think-aloud» på ratere ga ikke så fyldige data som empiri samlet inn via umiddelbare intervjuer.

I den femte fasen ble rammen for gjennomføring av intervjuene likt uavhengig av system. Intervjuguiden var todelt med en mer tilpasset del som ble benyttet innledningsvis. Dokumentasjonen fra prøvene var med for igjen å prøve å gjenerindre hendelsene og spesielt da det hadde gått noen dager. I denne fasen var vurdererne mer i refleksjonsfasen over de valgene de hadde tatt. Det finnes eksempler i intervjuene på at vurdererne tenker mer over sine valg og stiller også spørsmål til seg selv og sine bedømmelser.

Teorien det legges vekt på om «stimulated recall» er opptatt av å få fram det umiddelbare ved tankene til den intervjuede, og ulempene det er å få inn andre tanker og refleksjoner i intervjuet. Jeg har i mitt opplegg gitt større rom for at deltakerne kan slippe til med refleksjoner og mere «indirekte» tanker om vurderingsprosessen og derfor anvendt begrepet *stimulert oppmerksomhet*. Dette bunner i et ønske om større mangfold, noe som medfører et mer komplekst datamateriale. Det vil være data av ulik karakter, noe som kan medføre økt kompleksitet i analysen. Man kan se på dette som to datasett, hvor den ene er knyttet til delen med det umiddelbare intervjuet, og den andre er de data som kommer gjennom oppfølgingsspørsmål og generelle spørsmål.

Totalt sett er rammen og tenkningen bak forskningsdesignet knyttet til intervjudelen likt for alle tre systemene, men med behov for justeringer og tilpasninger. Empirien som framkommer totalt sett fremstår som fyldig og vil kunne bidra til å svare på mine forskningsspørsmål.



## 4.5 Dokumentanalyse

En mindre del av datainnsamlingen vil bestå av dokumentanalysen. Det vil hovedsakelig være knyttet til grunnlagsdokumenter som lover og forskrifter, læreplaner, dokumentasjon knyttet til de ulike vurderingssystemene og tilhørende oppgavetekster. Besvarelsene fra lærlinger vil ikke være en del av dokumentanalysen, men brukes som en støtte for mine informanter under intervjuene, og igjennom det bli tatt med i analysen gjennom vurderernes øyne.

Lynggaard (2015, s. 153) uttaler at empiriske undersøkelser er vanskelig å forestille seg uten dokumenter. De kan utgjøre et utgangspunkt for å utvikle undersøkelsesspørsmål. Det kan være knyttet en del metodiske utfordringer ved bruk av dokumenter i empiriske analyser. Lynggaard stiller spørsmålet om hva et dokument er. Det kan sies å være «... sprog som er fikseret i tekst og tid» (Lynggaard, 2015, s. 154). Han viser til begrepet tekst som i denne sammenheng kan favne bredt, hvor blant annet fotografi og teknologi inkluderes, og eksemplene på dokumenttyper er mange. Det anbefales å skille mellom det som er primære, sekundære og tertiære dokumenter. Primære dokumenter sees på som et dokument som et avgrenset antall mennesker får tilgang til. Dette er ikke rettet til den bredere offentlighet. Tilgangen til slike dokumenter vil kunne være begrenset og kan inneholde sensitive og følsomme opplysninger. Et sekundært dokument er et dokument som er offentlig tilgjengelig for alle når det måtte være. Et tertiært dokument er sett på som et dokument som alle har tilgang til og det karakteristiske ved dette er at det er produsert i etterkant av den begivenhet eller situasjon det refereres til. I praksis vil det ikke være vanntette skott mellom disse tre typer av dokumenter. Ser vi Lynggaards inndelinger av dokumenter opp mot HELVYRD-studien vil utvalget av dokumenter knyttes til alle de tre ulike kategoriene. Oppgavene i de ulike vurderingssystemene vil ses på som primærdokumenter. Det vil også gjelde besvarelsene til lærlingene og vurderingene til vurdererne. Det er i utgangspunktet et lite utvalg som skal ha tilgang til disse i starten, selv om det i ettertid vil bli offentlig for alle, og vil da bli sett på som sekundære dokument. Sekundære dokumenter i denne sammenheng vil være lover, forskrifter og læreplaner, samt sentrale dokumenter som sier noe om vurderingssystemene. Tertiære dokumenter vil være rapporter og oppsummeringen på de vurderinger som er gjennomført.

I forbindelse med dette sier Lynggaard at det er viktig å være klar over at forskjellige typer dokumentasjon gir forskjellige typer informasjon. Knyttet til HELVYRD-arbeid vil disse dokumentene gi en annen type informasjon eller data som enten styrker eller svekker de

konklusjonene jeg kan trekke ut av intervjudataene. Dokumenter kan ha en triangulerende (krysspeilende) funksjon på intervjudata og så vil «... et dokument likeledes kunne bidrage med data, som uttrykker det rationale, som dokumentet var uttrykk for på tidspunktet, hvor dokumentet ble produsert.» (Lynggaard, 2015, s. 156).

Et viktig element å være bevisst på er at et dokument ikke gir den direkte veien til sannheten om det aktuelle fenomenet. Hovedsaken er at data du får gjennom dokumentanalyse atskiller seg fra data du får igjennom intervjumetoder.

Søking i dokumenter kan følge «snøballmetoden» (s. 157) som stort sett handler om at man i en studie gjennom dokumenter følger de interne referanser helt til det ikke er noe mer å følge etter. I min studie vil det i mindre grad være behov for dette, men i samtalen med informanter har jeg vært nøye med å notere ned de dokumenter som de støtter seg til i sitt arbeide. Noen dokumenter, for eksempel opplæringsloven og læreplaner, kan være sentrale dokumenter (monumenter ifølge Lynggaard, s. 158) i min sammenheng siden de sier noe om rammevilkår for vurderingsarbeidet, men i min analyse vil disse kun være aktuelle som en forståelsesramme og en mulig referanse for informantene.

I analysen av dokumentene vil jeg følge en mer analytisk-induktiv undersøkelsesmetode (Lynggaard, 2015, s. 160), hvor jeg i liten grad vil foreta egne kildekritiske vurderinger. Mange av dokumentene det vil bli referert til er produsert under offentlig kjente omstendigheter. De framstår som autentiske, representative og med gitt innhold selv om vurderere, lærere, utdanningspolitikere vil kunne tolke dem forskjellig.

Som nevnt gjør jeg ikke en selvstendig dokumentanalyse, men bruker dokumentdata som en type metodetriangulering som bekrefter/avkrefter antakelser eller kaste nytt lys over temaene for min studie.

#### 4.6 Ethiske aspekter

Omsynet til ulike etiske aspekter ved designet er viktig for gjennomføring av datainnsamling. I alle faser av en undersøkelse oppstår etiske problemstillinger som vil prege intervjuundersøkelsen fra begynnelsen til slutt (Kvale & Brinkmann, 2015, s. 97; Patton, 2015, s. 495; Silverman, 2014, s. 141; Thagaard, 2013, s. 24). Kvale & Brinkmann (2015, s. 98) viser til behovet for å utvikle en etisk protokoll for å behandle spørsmål knyttet til ulike etiske aspekter som kan oppstå underveis. Det er viktig å være oppmerksom på hvordan det menneskelige samspillet påvirker intervjupersonenes atferd og selvforståelse (Kvale og

Brinkmann, 2015, s. 95). Kvale og Brinkmann (2015, s. 104) viser til flere etiske spørsmål som kan stilles rundt egen intervjuundersøkelse, summert opp i fire usikkerhetsområder. Samtidig vil jeg presisere at jeg forholder meg til de nasjonale retningslinjer om forskningsetikk (De nasjonale forskningsetiske komiteene, 2016)

Det første usikkerhetsområdet er *informert samtykke* (s. 104). For min del vil informert samtykke være knyttet til både MECVET-prosjektet og HELVYRD-prosjektet. De fire raterne knyttet til MECVET-testen har skrevet under på informert samtykke i forbindelse med MECVET- prosjektet som også gjelder for HELVYRD-prosjektet. Videre er det utviklet et eget informert samtykke til de kandidatene som er knyttet til de to norske vurderingssystemene (se vedlegg 2 NSD). Spørsmålet om hvor mye informasjon som bør gis og når er viktig å diskutere (Kvale og Brinkmann, 2015, s 105; Patton, 2015, s. 497; Silverman, 2014, s. 149; Thagaard, 2013, s 26). Det er viktig å være tydelig på hensikten med studien, og at dette blir klart for vurdererne. Jeg har valgt å få fram helheten og hovedmålet med studien, men gitt mulighet for muntlig gjennomgang og diskusjon av detaljer. I tillegg til de involverte vurdererne i selve studien har jeg vært nødt til å innhente tillatelse fra de institusjoner som er ansvarlige for de ulike vurderingene. Informasjonsbrevet har blitt sendt til de respektive, og jeg har fått skriftlige godkjenninger tilbake.

Det neste usikkerhetsområdet er *konfidensialitet*, som handler om å bli enige om hva dataene kan brukes til, noe som er et resultat av informantenes deltakelse. Viktige prinsipper er at private data som identifiserer deltakerne ikke skal avsløres, og at kvalitative data som samles inn, og skal inngå i offentlige dokumenter, ikke skal kunne direkte knyttes til informantene (Kvale og Brinkmann, 2015, s 106). Patton (2015, s. 496) stiller spørsmålet om hva som kan være fornuftige løfter for å sikre konfidensialiteten i prosjektet. Thagaard (2013, s. 23) viser til at de som gjøres til gjenstand for forskning har krav på at alt av informasjon om dem gjøres konfidensielt. Silverman (2014, s. 149) viser til viktigheten av at de som skal bli forsket på blir godt informert og får tilgang til det som kommer ut.

I MECVET og HELVYRD ble alle innsamlede data lagret på eksterne harddisker som har vært låst ned. Alle informantene fikk egne koder (krysstabell). Noe av utfordringen ved å velge ut prøvenemndsmedlemmer og sensorer er at det indirekte vil kunne være mulig å identifisere hvilke nemnder og sensorer som har deltatt. I disse fagmiljøene er ikke verden stor, det kan derfor være mulig å gjenkjenne hvor og hvilke nemnder og sensorer som har deltatt. Men det er igjen mindre sjans for å kunne identifisere den aktuelle informanten internt i de enkelte vurderingssystemene. Dette er det informert om i informasjonsskrivet.

Det tredje usikkerhetsområdet Kvale og Brinkmann (2015, s. 107) bringer på banen er *konsekvenser* som innebærer å ta hensyn til deltakere i den forstand at nytten kan være større enn ulempene for dem. Dette igjen vil gjøre undersøkelsen berettiget å gjennomføre.

Thagaard (2013, s. 30) viser til viktigheten av at forskeren skal respektere informantene, og det er viktig å tenke gjennom de konsekvenser dette kan ha for de involverte og sikre deres integritet. Patton (2015, s. 496) viser til en rekke spørsmål, blant annet om det finnes ulike faktorer som kan sette informanten i en risikofylt situasjon. I min studie vil det ikke være overhengende fare for at noen vil komme i en situasjon hvor de føler seg ukomfortable eller truet. Bakgrunnen for dette er at de aller fleste som deltok i min studie er folk med lang erfaring innen vurdering. Kvale og Brinkmann (2015, s. 107) viser til behovet for å kjenne til undersøkelsesfeltet for i større grad kjenne til hva man skal gjøre. Jeg har relativt god kjennskap til dette feltet.

Når det gjelder *forskerens rolle* peker Kvale og Brinkmann (2015, s. 108) på betydningen av integriteten til forskeren og avgjørende faktorer knyttet til kunnskap, erfaring, ærlighet og rettferdighet. Dette igjen er en avgjørende faktor for den kvalitet man får på den vitenskapelige kunnskap, samt de etiske beslutninger som treffes i kvalitativ forskning. Vitenskapeligheten på den kunnskapen og de funn som legges fram og offentliggjøres må være så nøyaktig og representativt for forskningsområdet som mulig. Validering og kontroll samt gjennomsiktighet av rapporter er grunnlaget for de konklusjoner som har blitt tatt er viktig å etterstrebe. Et siste moment er forskerens rolle og personlige samspill med intervjupersonene, og hvordan dette kan påvirke forskeren. Det er viktig å kombinere nærhet med avstand (Kvale og Brinkmann, 2015, s. 108) og Thagaard (2013, s. 74) mener det er viktig å tenke over hvordan man skal presentere seg i det miljøet man skal forske. Patton (2015, s. 429) trekker inn «... *overcoming the overconfidence of incompetence*» (Patton 2015, s. 429). Han henviser til flere studier som viser at mennesker som utfører ulike handlinger dårlig er samtidig veldig trygge på tilstrekkeligheten av egen kompetanse og de er tryggere enn de som faktisk utgjør handlingene bra. Patton bringer også på banen spørsmålet om den mentale helsen til intervjueren og hvordan man takler og behandler den informasjonen man får underveis og hvordan dette blir i ivarett. Kvale og Brinkmann (2015, s. 108) viser til at det viktigste forskningsinstrumentet er forskeren selv som person. Forskerens kunnskap, ærlighet, rettferdighet og integritet blir avgjørende for kvaliteten på den kunnskap som produseres. Man setter ikke sin lit utelukkende til personlig intuisjon, eller inntar den holdning at alt er mulig (intellektuelt håndverk).

Et viktig moment for meg som forsker med min bakgrunn i feltet er å være bevisst min forutinntatthet og forforståelse basert på mine egne meninger om vurderinger. Kvernbekk (2005, s. 18) definerer en insider som «*en person som på grunn av sin spesielle og privilegerte posisjon innehar – eller antas å inneha – en kunnskap som andre ikke har*» (Kvernbekk, 2005, s. 18). Med dette som utgangspunkt vil jeg være en insider som både har kunnskap om og praktisk erfaring med det feltet jeg skal forske på. Min forsker-rolle kan sees på som todelt. Jeg har deltatt som forsker i MECVET-prosjektet samt mitt hovedansvar som forsker i HELVYRD-prosjekt. Arbeidet i MECVET har medført at jeg har deltatt på arbeid som har hatt direkte og indirekte betydning for mitt eget arbeid i HELVYRD. Jeg mener at min integritet ivaretas med min gode kjennskap til feltet, og mener jeg har et godt grunnlag for å stille relevante spørsmål selv om det konkrete fagfeltet ikke er mitt.

#### 4.7 Koding av empiri og utvikling av kategorier

Analysing av data begynner allerede i intervjusituasjonen (Kvale og Brinkmann, 2015, s. 221). Ved de innledende intervjuene og i oppfølgingsintervjuet ble det gjort notater om vurdererens tanker om de vurderinger de gjennomførte. Denne prosessen inkluderte en gjennomhøring av lydbåndene fra førsterunde intervjuene i arbeidet med å utvikle oppfølgings spørsmål til andre runde intervjuer. Dette ble en naturlig del av analyseprosessen. Deretter ble arbeidet med transkribering av lydfiler en naturlig del av analyseprosessen. Først for intervjuene gjennomført i januar og februar 2015 knyttet til fagprøven og MECVET-test (ferdig transskribert i april), og i juni 2015 for eksamen. Alle intervjuer var ferdig transkribert til 1. juli 2015 (178 A4-sider). Under transkriberingen foregikk det analyser fortløpende, og enkeltnotater ble gjort som utgjorde et bidrag i utviklingen koder og kategorier (Kvale og Brinkmann, 2015, s. 218). Ved å gjøre transkriberingsjobben selv får man bedre kjennskap til datamaterialet, samt at jeg får innblikk i min egen intervjustil (Kvale og Brinkmann, 2015, s. 207). Jeg har i størst mulig grad lagt vekt på å gjengi ordrett det som ble sagt. Ønsket var å få en mest mulig lettlest utgivelse fra de enkelte historiene. Kvale & Brinkmann (2015, s. 211) viser til transkripsjonens reliabilitet og validitet. Jeg har i min transkripsjon lyttet til lydopptakene både to og tre ganger noe som jeg mener gir pålitelighet i min tolkning av det informantene uttrykker. Enkelte deler var utydelig, men samtidig husker man også i hvilken kontekst dette ble sagt og det hjelper i forståelsen av tolkningen. Validiteten kan settes på prøve i den forstand at man kan tolke et utsagn mer i retning av hva man legger i det selv enn det informanten mente. Det er viktig i størst mulig grad med objektivitet å ikke legge i det hva

man selv kanskje har som en forforståelse. Etisk er det viktig med anonymiteten og sikre konfidensialiteten til informantene, slik at det i størst mulig grad blir vanskelig å kunne spore.

I prosessen rundt koding og kategorisering har jeg bygget på Charmaz (2014, s. 109) sin koding- og kategoriseringsprosess forankret i Grounded theory, og redegjørelsen fra Johannessen, Christoffersen og Tufte (2016, s. 181), støttet av Thagaard (2013, s. 181) sine tanker og prosess om tematisk analyse. Jeg vil understreke at denne studien ikke skal sees på som en ren grounded theory eller tematisk analyse.

I arbeidet med koding og kategoriseringsprosess har jeg benyttet dataanalyseverktøyet Atlas TI. Deler av funksjonaliteten til programmet er benyttet som opplasting av de transkriberte tekstfilene samt kodingsprosessene og utkjøring av filer. Dette har bidratt i første omgang til utviklingen av kategorier og underkategorier. De eksporterte filene med koder er gjort i et Excel regneark, hvor en videre bearbeiding i systematiseringen av koder og kategorier vil skje. «Penn & papir» er også benyttet i arbeidet for utvikling av skisser til å se helheter og sammenhenger i datamaterialet.

Etter at filene med transkriberte intervjuer ble lastet inn i Atlas TI, har disse blitt gått igjennom med koding og lytting av lydbåndfilene parallelt. Charmaz (2014, s. 111) uttaler at koding betyr å navngi segmenter med data, og stille spørsmål om hva du kan få ut av disse. Charmaz beskriver to hovedfaser. Den første er *initial* fase (s. 109) som omhandler en innledende koding ved å navngi ord, linjer eller segmenter av data. Den andre er *focused* fase (s. 138), som bygger på de mest nyttige eller hyppigste *initial* kodene. Disse oppstår oftere og har mer signifikans enn andre koder. Dette er med på å kondensere og spisse hva som ble gjort i *initial* koding, og får fram hva som er essensielt i empirien som skal analyseres. Thagaard (2013, s. 182) sier at temasentrerte tilnærminger handler om å knytte kodeord til utsnitt av data som sier noe om det meningsinnholdet som ligger i den aktuelle teksten. Ut i fra dette vil de utsnitt som omhandler det samme tema være grunnlaget for å utvikle kategorier.

Ved *initial* koding har strategien vært å gå inn i empirien med åpne øyne for å prøve å forstå de intervjuedes meninger og situasjoner og de handlinger de gjør. Charmaz uttaler: «*Try to remain open to seeing what you can learn while coding and where it can take you*» (Charmaz, 2014, s. 117). Samtidig har det i tidligere faser (intervju, transkribering og gjennomgang av litteratur og tidligere forskning) fått innspill som trekkes med inn i analyseprosessen. I tillegg

kan deler av spørsmålene fra intervjuguiden, spesielt de generelle, inneholde begreper og fokusområder som ubevisst kan utgjøre koder. Med tanke på forhåndsbestemte koder eller områder/temaer har det blitt mer fokus på dette (Charmaz, 2014, s. 114; Thagaard, 2013, s. 181?), og i den sammenheng uttrykker Charmaz: «*Make your codes fit the data you have rather than forcing the data to fit them*» (Charmaz, 2014, s. 120). I kodingsprosessen viser Charmaz (2014, s. 124-132) til ulike former for koding som om handler «word by word», «line by line» eller «incident by incident». I min koding har jeg en blanding av disse ulike variantene, men mest brukt de to sistnevnte. Figur 26 viser et eksempel på kodingsprosessen ved bruk av Atlas TI.



Figur 26. Et eksempel (utsnitt) fra ratere i MECVET på hvordan kodingen har foregått ved bruk av Atlas TI.

Etter den første runden ble materialet liggende en periode før en ny runde ble gjort. All kodingen ble på nytt gjennomgått med lytting av lydfilene parallelt, og det ble gjort rydding og omkoding. Charmaz (2014, s. 118) sier at det er en naturlig del i prosessen å oppdage hull og gap i dataene. I HELVYRD-studien kunne dette delvis bli fanget opp gjennom å ha hatt to runder med intervjuer. Samtidig er det ikke mulig å kunne dekke alt, og dette vil kunne være til gjenstand for senere forskning på feltet.

Om *focused* koding uttaler Charmaz (2014, s. 138) at det handler om å studere og vurdere *initial* koding. Overgangen fra *intial* til *focused* koding er ofte sømløs, hvor de mest synlige *intial* koder som i tillegg har en viss teoretisk retning er mer sentrale og blir med videre i analysen. Et annet interessant element er at med de kodene man har, kan man i analysen avdekke fenomener som mange har erfart, men som ikke er konseptualisert. Som en del av *focused* koding var den videre systematiseringen av de eksporterte kodene fra Atlas TI inn i Excel regneark vesentlig som en støtte i arbeidet. I fasen med *focused* koding vil kodene bli mer konseptuelle enn de ulike *intial*-kodene Charmaz (2014, s. 138). Samtidig skjer *initial* og *focused* koding i fremvoksende prosesser. Det handler om å gjøre valg, at disse er tentative, og at det er fleksibilitet til å endre og flytte på disse.

Figur 27 viser et utsnitt av de koder (76 totalt (alfabetisk)) som ble utviklet i første runde og eksport fra Atlas TI. Fordelingen vises i de forskjellige transkriberte dokumentene (P10-P13 Eksamen; P14-P16 Fagprøven; P17-P18 MECVET-test).

CODES-PRIMARY-DOCUMENTS-TABLE										
Report created by Super - 26.11.2015 15:18:57										
HU: [H:\AY\AA-PhD-MECVET\PhD\PhD-avhandlingen HELVYRD\PhD monografiavhandling...\Intervjudata.hpr7]										
Code-Filter: All [76]										
PD-Filter: All [9]										
Quotation-Filter: All [814]										
	P10: Eksa	P11: Eksa	P12: Eksa	P13: Eksa	P14: Fag	P15: Fag	P16: Fag	P17: MEC	P18: MEC	TOTALS:
Besvarelse Løsninger og innspill fra kan	1	1	2	2	0	0	1	0	0	7
Besvarelse og fokuserer på	1	5	1	1	2	1	10	2	8	31
Besvarelse og mening	0	1	0	3	2	0	1	4	11	22
Besvarelse ulike løsninger	0	1	0	1	0	0	1	2	3	8
Besvarelse utdyping bra	1	0	4	4	6	4	9	11	9	48
Besvarelse utdyping mangler	2	2	3	7	5	2	11	13	8	53
Bilder i oppgavetekst	1	2	0	0	0	0	0	0	0	3

Figur 27. Et utsnitt av initial koder og fordeling i de ulike vurderingssystemene

Bilde viser hyppigheten og fordelingen av kodingene på et gitt stadium i analysen og hvordan denne tabellen kan støtte den videre «focused» koding. I den neste fasen ble det gjort ett par runder med revideringer av koder samt en sorteringsjobb som knytter kodene direkte til de ulike vurderingssystemene, og så samlet jeg de kodene som var felles for alle systemene



(figur 28, et utsnitt).

	P10: Ek:	P11: Ek:	P12: Ek:	P13: Ek:	P14: Fag:	P15: Fag:	P16: Fag:	P17: MEC:	P18: MEC:	TOTAL
Diskusjonspartner	0	2	0	0	0	0	0	0	0	2
Eksamensform uttalelse negativ	0	1	1	0	0	0	0	0	0	2
Eksamensform uttalelse positiv	1	0	1	2	0	0	0	0	0	4
Eksamensformen	1	6	1	1	0	0	0	0	0	9
Helhetlig arbeid	0	1	0	0	0	0	0	0	0	1
Karakter eksamen 1-6 redegjørelse	1	3	1	1	0	0	0	0	0	6
Karakter eksamen begrunnelse	1	3	0	6	0	0	0	0	0	10
Karakter eksamen Lav, middels og høy	1	3	0	1	0	0	0	0	0	5
Karakter eksamen og tvill	3	1	3	5	0	0	0	0	0	12
Karakterspenn eksamen	0	1	0	0	0	0	0	0	0	1
Sensor uenighet	3	0	3	4	0	0	0	0	1	11
Vurderingsgrunnlaget ift kriterier	0	1	0	0	0	0	0	0	0	1
Vurderings tvill	1	1	1	1	0	0	1	0	0	5
Vurderingskriterier Eksamen	1	4	1	1	0	0	0	0	0	7
Egenvurdering uttalelse	0	0	2	0	0	3	0	0	0	5
Fagprøveformen	0	0	0	0	3	1	8	0	0	12
Karakter fagprøve begrunnelse	0	0	0	0	8	10	7	0	0	25
Karakter fagprøve 1-B-M redegjørelse	0	0	0	0	5	5	6	0	0	16
Karakter fagprøve tvill	0	0	0	0	1	6	4	0	0	11
Nemdsmedlem og rolle	0	0	0	0	1	3	0	0	0	4
Nemdsmed uenighet	0	0	0	0	2	3	2	0	0	7
Nemdsmed uenighet	0	0	0	0	2	0	1	0	0	3
Vurderingskriterier Fagprøve	0	0	0	0	11	7	10	0	0	28
Løsningsrommet MECVET	0	0	0	0	0	0	0	2	2	4
MECVETtestformen	0	0	0	0	0	0	0	7	12	19
Nivågradering ikke oppfylt- oppfylt M	0	0	0	0	0	0	0	2	4	6
Nivåkriterier MECVET	0	0	0	0	0	0	0	18	13	31
Rater enighet	0	0	0	0	0	0	0	4	3	7
Rater uenighet	0	0	0	0	0	0	0	8	18	26
Rating og tvill	0	0	0	0	0	0	0	5	4	9
Tolkning MECVET kriterier	0	0	0	0	0	0	0	1	0	1
Vurderingskriterier MECVET	0	0	0	0	0	0	0	3	9	12
Besvarelse Løsninger og Innspill fra ka	1	1	2	2	0	0	1	0	0	7
Besvarelse og fokuserer på	1	5	1	1	2	1	10	2	8	31
Besvarelse og mening	0	1	0	3	2	0	1	4	11	22
Besvarelse ulike løsninger	0	1	0	1	0	0	1	2	3	8
Besvarelse utdyping bra	1	0	4	4	6	4	9	11	9	48
Besvarelse utdyping mangler	2	2	3	7	5	2	11	13	8	53
Helhetlig vurdering	1	4	5	3	5	2	5	2	3	30
Sammenligning av kandidater	0	0	0	0	1	3	2	0	2	8
Solusjonsdeltaker	0	0	0	0	0	0	0	0	0	0

Figur 28. Utsnitt av sorterte koder knyttet til vurderingssystem

Neste fase i prosessen handler om å utvikle en kategorisering. Det ble i første omgang utviklet kategorier med underkategorier som i hovedsak ble utviklet basert på kodene beskrevet over. I prosessen ble noen koder til en underkategori, samt at det er koder som har blitt slått sammen og dannet en ny underkategori. Figur 29 viser en påstartet prosess med kategorisering i hvert vurderingssystem. Det vesentlige er at de utviklede kategorier og underkategorier i stor grad er parallele i vårt vurderingssystem.

Vurderingssystemer					
<b>Eksamen</b>		<b>Fagprøve</b>		<b>MECVET</b>	
<b>Besvarelse</b>	Fokus på	<b>Besvarelse</b>	Fokus på	<b>Besvarelse</b>	Fokus på
	Mening		Mening		Mening
	Utdyping bra		Utdyping bra		Utdyping bra
	Utdyping mangler		Utdyping mangler		Utdyping mangler
	Begrunnelse V		Begrunnelse V		Begrunnelse V
			Viktigste element V		
	Ulike løsninger		Ulike løsninger		Ulike løsninger
	Innspill fra kand		Innspill fra kand		Innspill fra kand
<b>Sensor</b>	Enighet	<b>Nemndsmed</b>	Enighet	<b>Rater</b>	Enighet
	Uenighet		Uenighet		Uenighet
			Rolle		
<b>Vurdering</b>	Helhetlig vurdering	<b>Vurdering</b>	Helhetlig vurdering	<b>Vurdering</b>	Helhetlig vurdering
	Vurderingsprosessen		Vurderingsprosessen		Vurderingsprosessen
	Sammenligning??		Sammenligning		Sammenligning
	Fagperson		Fagperson		Fagperson
	Reellt yrke		Reellt yrke		Reellt yrke
	Vurd- kriterier		Vurd-kriterier		Vurd-kriterier
					Løsningsrommet
	F vs E		F vs E		M vs F
					E vs M
<b>Karakter</b>	Redegjør 1-6	<b>Karakter</b>	Redegjør IBM	<b>Karakter</b>	Redegjør Nivå
	L-M-H		Begrunnelse		Nivåkriter
			Tvil		Tvil
<b>Formen</b>	Pos	<b>Formen</b>		<b>Formen</b>	
	Neg				
<b>Oppgaver</b>	Oppgaver og tekst	<b>Oppgaver</b>		<b>Oppgaver</b>	Oppgaver og tekst
	Meninger oppgave		Meninger oppgave		Meninger oppgave
	Bilder				
	PC bruk		PC bruk		PC bruk
	Skriving		Skriving		
	Språk				
	Tegninger		Tegninger		Tegninger
<b>Kompetanse</b>	Kompmål og læreplan	<b>Kompetanse</b>	Kompmål og læreplan	<b>Kompetanse</b>	Kompmål og læreplan
	På papiret				Kompmål og dimensj
	Læring		Læring		
			Prak/Teori		Prak/Teori
<b>Faglig</b>	Fagarbeider redegjør	<b>Faglig</b>		<b>Faglig</b>	Fagarbeider redegjør
	Ingeniør/Fagarbeider		Ingeniør/Fagarbeider		Ingeniør/Fagarbeider
	Jobbprosess				Jobbprosess

Figur 29. Et av de første utkastene til kategorisering av koder.

Thagaard (2013, s. 182) uttrykker at man studerer og analyserer ulike temaer som har blitt utviklet gjennom bearbeiding av datamaterialet fra intervjuene. I prinsippet skal data fra hver deltaker bidra til å utvikle temaene. I HELVYRD vil dette variere noe på grunn av vurderingssystemenes form. Når det gjelder antall kategorier, bør det være et passende antall. Med for mange kategorier kan sammenligningen bli uoversiktlig, og med for få vil nyansene i analysen kunne bli for lite. Ved kategoriene utviklet i HELVYRD-prosjektet vil det være sammenlignbart, men ved noen underkategorier vil det ikke være mulig grunnet vurderingsformens form. I de tilfellene vil kun ett av systemene være representert.

Omfanget av mitt datamateriale gir muligheter for å gå analytisk til verks på større mengder av mitt materiale for koding og kategoriutvikling. Ved å trekke inn store deler av empirien innledningsvis i analysen mener jeg det i større grad vil gi en helhet og sammenheng som kan gi mer innsikt i problemfeltet som sådan. Totalutviklingen av kategorier for datamaterialet har skjedd med både induktive og deduktive innfallsvinkler. Det vil si at jeg har en posisjon mellom disse to som utgjør et samspill, altså en abduksjon (Thagaard, 2013, s. 201). Noe utvikling av teori skjer igjennom analysene av empirien mens andre deler av empiri blir knyttet til allerede utviklede begreper fra teori eller andre studier. Dette igjen gir muligheter til å trekke linjer fra egen forskning til tilsvarende fenomener i andre studier. Min egen erfaring innen fagfeltet vil også kunne bli en del av prosessen, spesielt da med en mer induktiv innfallsvinkel.

*Axial* koding handler om å relatere kategorier til underkategorier og se nærmere på egenskapene og dimensjonene i en kategori (Strauss & Corbin, 1998, s. 124). Det er ønskelig å bringe alle fragmenter av data igjennom kodingsprosessen sammen igjen til en koherent helhet. Relasjonene mellom kategorier oppstår på ett konseptuelt nivå mer enn et deskriptivt nivå (s. 125). Charmaz uttaler: «*Although axial coding may help researchers to explore their data, it encourages them to apply an analytic frame to the data*» (Charmaz, 2014, s. 149). *Axial* koding kan begrense hva forskeren får ut av studien ved at kodene som blir konstruert rammes inn. *Axial* koding kan oppleves som klargjørende i analysen av den fremvoksende empirien, men samtidig kan det bli for teknifisert, og føre til en mer overfladisk analyse og konklusjon. Det er vesentlig for meg å være bevisst på dette i eget arbeid i forhold til den måten kodingen og kategorisering har blitt gjennomført. Thagaard (2013, s. 188) bringer dette på banen som temasentrerte tilnærminger og viktigheten av at konklusjoner kan framheve sammenhenger mellom de temaer og kategorier som er representert. Det kan gi en mer helhetlig forståelse av vurderernes vurderinger i ulike kontekster og systemer. Spørsmålet jeg stiller er om det kan bli for teknisk med både bruk av Atlas TI og Excel regneark, eller om det er mulig å holde fokus på vesentlige funn i min empiri? Kritikken Thagaard (2013, s. 191) peker på er rettet inn mot at det velges bare enkelte avsnitt ut fra helheten som en tekst utgjør av en intervjuet person. Det er viktig å være bevisst på hvilken del av sammenhengen denne teksten tilhører, og at dette vurderes opp mot den konteksten der selve samhandlingen har foregått. Gjennom å analysere sammenhengen av de utvalgte temaene vil en helhetlig forståelse av materialet utvikles. I arbeidet med kondensering av empiri og tekst som er

identifisert som meningsbærende er aktuelle kategorier (Johannessen et al., 2016, s. 176) med på å belyse mine forskningsspørsmål.

De foreløpige empiriske kategorier og underkategorier ble så systematisert opp mot de teoretiske aspektene fra kapittel 3 (kalt «teoretiske kategorier» i tabellen og som vil bli kalt tema i kapittel 5, inspirert av Thagaard sin tematisering) (figur 30) presentert i kapittel 3.1 (3.1.3-3.1.8).

Revidert utgave					
Teoretiske kategorier					
Formål og intensjon	Mål og kriterier	Rammer og kontekst	Metode og verktøy	Innhold	Forutsetninger
Empiriske katgo m/ukat					
Vurdering	Kompetanse	Formen	Vurdering	Besvarelse	Vurdering
Helhetlig vurdering	Kompmål og læreplan	Form og struktur vurd-sys	Vurderingsprosessen	Fokus og Mening	Fagperson
	Kompmål og dimensj M	FvsE MvsF EvsM	Sammenligning??	Utdyping bra m/beg	Reellt yrke
	Vurd- kriterier/løsningsrom			Utdyping mangler m/beg	
				Ulike løsninger/inns spill	
	På papiret?				
Vurderer	Faglig	Oppgaver			
Samstemthet og tvil	Faglige redegjørelser fagar/ing	Oppgaver og tekst og meninger			
		PC bruk,skriv, tegning,språk,bilde			
	Karakter				
	1-6 LMH, IBM,Nivå				

Figur 30. En skisse på sammenhengen med teoretiske perspektiver og empiriske kategorier og underkategorier

Som et siste grep i prosessen med systematisering av tema og de respektive kategorier basert på grunnlaget presentert i figur 30, har underkategoriene i den endelige fremstilling blitt omstrukturert til overordnede kategorier hvor enkelte ble slått sammen og andre er blitt stående. Koblingen til de enkelte tema ble hovedsakelig beholdt, unntatt tema formål og intensjon hvor de tilhørende kategorier og underkategorier ble omplassert eller systematisert inn i de andre kategoriene (se figur 31).

Tema	Kategori	Områder
Innhold	Vurderernes fokus	11 innholdselementer
Rammer og kontekstuelle faktorer	Vurderingssystemenes form og sammenheng  Utforming av oppgaver og tekst	Formen Reelle gjennomføring Sammenheng Utforming Åpenhet
Mål og kriterier	Kompetanse og faglige redegjørelser Vurderingskriterier  Karakterens betydning	Forholde seg til Ulike nivå Benytte kriteriene Stramhet -----
Metoder	Vurderingsstrategier  Samstemthet	Planlegging Fremgangsmåter Sammenligninger Enighet Uenighet

Forutsetninger	Fagperson som vurderer	Faglig bakgrunn Skole og arbeidsliv Yrkeslivet
----------------	------------------------	--

Figur 31. Forholdet mellom tema (koblet til teoretiske aspekter om helhetlig vurdering) og empiriske kategorier samt områder for de tre vurderingssystemene.

Til de respektive kategoriene vil det i presentasjonen framkomme sentrale *områder* i empirien. Disse er et resultat av enkelte kategorier som er endt som områder, samt at i prosessen med å skrive ut resultater så har disse områdene fremkommet som naturlige å ha med i presentasjonen.

Tabellen er et resultat av en lengre prosess, og har ikke vært et kronologisk forløp, men et arbeid som har gått fram og tilbake under arbeidet med analyse, fortolkninger og diskusjoner. Tabellen gjenspeiler et endelig arbeid som har gitt en systematisering som har vært naturlig for denne studien. Ikke sagt at det ikke kunne vært gjort annerledes både prosessen og sluttresultatet. Siden studien er av en eksplorerende karakter, så har jeg valgt å gå noe bredt ut, noe som kan gjenspeiles i temaer, kategorier og områder.

Skal kort redegjøre for overgangen fra ovenstående gjennomgang av analyse og utvikling av tema og kateogrier til kapittel 5 og delvis kapittel 6. Under hvert tema i kapittel 5 blir det en presentasjon og diskusjon av resultater fra alle vurderingssystemene i tilknytning til hver kategori. Sitater fra de intervjuede merkes deres kode samt nummer på dokument hvor rådata er lagret, for eksempel E3-P11. Videre under hvert tema vil det bli et kapittel med diskusjon av presenterte resultater hvor tilhørende teori om aspekt ved helhetlig vurdering og tidligere forskning trekkes inn. Så følger et kapittel som presenterer funn i hvert tema. Funn vil bli tatt med i en sammenfattende diskusjon i kapittel 6. Redegjørelsen om gjennomføring vil bli tatt opp igjen innledningsvis i de respektive kapitlene.

Det kan oppleves å være en overlapping mellom kategoriene siden det er en naturlig relasjon mellom disse. De er ikke isolerte enheter. Som et resultat av det vil det kunne oppleves at noe av den analyserte empirien kunne vært i plassert i en annen kategori. I den sammenheng vil det for denne studien være vesentlig å se på sammenhenger og helheter mellom temaene, noe som vil styrke forskningens fokus, samt aksial koding.

#### 4.8 Oppsummerende betraktninger om kvaliteten og validitet på design og metode

Som en avrunding vil jeg gjøre noen betraktninger om kvaliteten ved å støtte meg til Tracy (2010, s. 837) sine «Eight «Big-Tent» Criteria for Excellent Qualitative Research». Tracy har

prøvd å konseptualisere åtte punkter for å utvikle en pedagogisk modell som skal gi et felles grunnlag å diskutere «best practices» innen kvalitativ forskning. Jeg vil kort redegjøre for punktene og fortløpende diskutere opp imot eget forskningsopplegg. Som en ytterligere støtte vil studiens validitet (diskutert i kapittel 4.3.1, 4.4.4 og 4.7) diskuteres ytterligere ved å anvende Kvale og Brinkmann (2015, s. 278) sine syv stadier av validering (tematisering, planlegging, intervjuing, transkribering, analysing, validering og rapportering).

Det første punktet til Tracy handler om «Worthy topic», om det valgte temaet er relevant, i rett tid, signifikant og interessant. Er det «worthy» vil dette være et nyttig tilskudd til forskningsfeltet, og vil engasjere de som leser og hører om dette. Kvale og Brinkmann viser til *tematiseringen* og hvor godt undersøkelsen er forankret i teorien og gyldigheten til de utlede forskningsspørsmålene. I HELVYRD er fokuset på vurdering innen fag- og yrkesopplæring et viktig område, og er lite forsket på. Studien vil være et bidrag til eksisterende studier, samt sette fokus på områder som det må forskes mer på. Ett av elementene som er interessant ved denne studien er at flere vurderingssystemer er trukket inn mot det samme faget, noe som kan gi en utvidet forståelse for betydningen av de ulike vurderinger.

Punkt to handler om «Rich rigor», som fokuserer på å kunne belyse et komplekst felt, noe som også krever et mangfoldig metodisk verktøy for gjennomføring. Det gjelder tilgangen på teori, datamengden, tiden, benyttede prosedyrer i feltet, utvalgte informanter og analyse av innsamlede data. Samtidig pekes det på at et hode fullt av teori og data ikke nødvendigvis vil gi den beste kvalitative studien. Relatert til mengde av data er det viktigste om studien kan «... *provide for and substantiate meaningful and significant claims*» (Tracy, 2010, s. 841). Kvale og Brinkmann viser til *planlegging* med fokus på kvalitet av undersøkelsesopplegget og de metoder som anvendes. Fokuset i HELVYRD-studien er mangesidig fordi det å inneha en yrkeskompetanse som skal vurderes er så komplekst. Dette har medført en datainnsamling som er relativt omfattende og kompleks i seg selv, hvor målet har vært å være så nær feltet som mulig. En triangulering med kvalitative data gjennom intervjuer og dokumentanalyser, samt noe kvantitative data fra MECVET, gir et mer komplekst datasett som grunnlag for å besvare de aktuelle forskningsspørsmålene.

Det tredje punktet handler om «Sincerity» og innebærer selv-refleksjon med fokus på egne verdier, bias og tilbøyeligheter. Det er vesentlig med transparens ved metoder og utfordringer. Kvale og Brinkmann peker her på hvilke *valideringsformer* som er relevante hvor blant annet

hva som kan være et egent forum for dialog om resultatenes gyldighet. I HELVYRD har jeg hele tiden søkt å være bevisst min egen rolle som forsker, og å være åpen og ærlig om styrker og svakheter. Det har vært viktig å være bevisst min egen faglige bakgrunn og mulighet for forutinntattheten i forskningen. Mitt ønske er å være så transparent som mulig, slik at lesere får et best mulig bilde av hvordan forskningen har foregått. Jeg opplever selv utfordringer med metodene for datainnsamling. De har vært krevende for meg, samtidig har det gitt mulighet for å skaffe empiri som jeg mener er valide og reliable. Transkriberingen kunne vært gjort mer nøyaktig, og i starten fulgt mer standarder for gjennomføring. Dette har ført til merarbeid.

Det fjerde punktet, «Credibility» fokuserer på «thick descriptions», noe som innebærer detaljert beskrivelse av hendelser i den aktuelle kontekst, å få tak i den tause kunnskapen, å krystallisere og triangulere. «Multivocality» innebærer ha flere stemmer inn, og handler om å vise fremfor å fortelle. «Member reflections» inviterer deltakere inn i analyseprosessen og produksjonen av rapporten. Kvale og Brinkmann peker på to ting som kan knyttes til dette. Det første er *intervjuing* som omhandler intervjueren og hvordan intervjuet foregår. Det andre er *analyseringen* er hvordan man forholder seg til intervjuteksten og at tolkningen er logiske. I prosessen rundt HELVYRD er det søkt å samle inn nok data med ulike metoder. En triangulering er gjort gjennom flere faser med intervjuer og dokumentanalyse, mens en mer omfattende krystallisering kunne vært jobbet mer med. Den tause kunnskapen hos informantene krever mye tilstedeværelse, men med de metodene som er brukt mener jeg det har vært mulig å komme nærmere inn på de tanker som den enkelte har om sine vurderingsprosesser. Når det kommer til analysen, har jeg stort sett forholdt meg til datamaterialet alene. Det har ikke vært rom for å involvere informantene i så stor grad. Men det å gjennomføre to runder med intervjuer har gitt mulighet til å diskutere det som ble sagt i første runde. I HELVYRD-studien ville det vært interessant å involvere informantene i etterkant for å diskutere de erfaringer som er gjort. Jeg er ikke i tvil om at det ville berike datamaterialet ytterligere.

Femte punktet handler om «resonance» som innebærer «aesthetic merit» og «generalizability og transferability». Det første handler om å presentere en tekst i en «... *beautiful, evocative, and artistic way*» (Tracy, 2010, s. 845). Det oppfordres til å skrive på en kreativ og kompleks måte som engasjerer leseren, og som kan «... *surprises, delights, and tickles smething within us*» (Tracy, 2010, s. 845). Det andre handler om at studien kan ha verdi på tvers av ulike kontekster og situasjoner. Kvale og Brinkmann viser til *rapportering* hvor det skal gis en

valid beskrivelse av hovedfunnene i studien. I HELVYRD har jeg lagt vekt på å få fram en struktur som gjør at avhandlingen henger mest mulig sammen og som er lett for en leser å følge. Om det engasjerer eller overrasker, får være opp til leseren selv, men jeg har lagt vekt på å være så nøyaktig og informativ som mulig uten å prøve å lage så mye ekstra rundt presentasjonen. Det er nok i min natur å være systematisk og rett på sak. Når det gjelder generalisering og overføringsverdi, mener jeg studien bidrar til diskusjonen om generalisering av funnene, og at dette også kan trekkes inn i andre situasjoner. Larsson (2008, s. 33) redegjør for fem ulike retninger om synet på generalisering. Et av de omhandler «*Generalization through recognition of patterns*» som betyr at man gjennom å sammenligne og sortere data kommer fram til noen kjerne kategorier og sammenhengen mellom disse. Bare det at studien i seg selv handler om tre vurderingssystemer, gjør at generalisering og overføring allerede diskuteres i studien.

Det sjette punktet handler om «Significant Contribution» til det teoretiske feltet, som har heuristisk signifikans, og at den utviklede kunnskapen i forskningen er anvendelig i det praktiske feltet. Det kan vi også knytte til Kvale og Brinkmann sitt punkt om *validering* og resultatenes gyldighet. I HELVYRD er kunnskapen som har framkommet relevant, spesielt siden feltet er lite forsket på. Funnene i denne studien bekrefter allerede tidligere forskning, samtidig som det gir grobunn for mer diskusjon og videre forskning. Studien gir refleksjoner som vil være anvendelige i den praktiske gjennomføringen av de ulike vurderingsseansene.

Det syvende punktet handler om «Ethical» forskning og innebærer «procedural ethics», hvor blant annet de data som samles inn blir godt og trygt bevart på et sted ingen andre har tilgang. «Situational ethics» innebærer at hver hendelse er forskjellig, noe som betyr at forskeren må reflektere over, kritisere og stille spørsmål om etiske valg. Forskeren skal konstant reflektere over metoden og de data som skal presenteres. Kvale og Brinkmann viser til at man *planlegger* for at man produserer «... kunnskap som er fordelaktig for mennesket og minimaliserer skadelige konsekvenser» (s.278). I HELVYRD har jeg så langt som mulig etterstrebet å være etisk i mine handlinger rundt forskningen (også diskutert i kapittel 4.5). Jeg har hatt flere runder med spørsmål om ulike sider av arbeidet, og vært nøye på å opplyse mine informanter om hensikten med studien.

Det siste og åttende punktet handler om «Meaningful Coherence» og viser til at studien må nå sine mål, bruker metoder som er forenlige med de mål som skal nås, og på en meningsfull måte koble sammen litteratur, forskningsspørsmål, funn og tolkninger. Dette punktet kan også



knyttet til Kvale og Brinkmann sitt stadiet om rapportering. I HELVYRD har det vært utfordrende å koble alle deler, slik at mine mål blir nådd og presentert på en tydelig måte. Spesielt har det vært utfordrende å få til et teoretisk rammeverk som skal henge sammen, og som skal brukes i analysene. Strukturen på presentasjon, analyse og sammenfattende diskusjon har krevd mange runder med tenkning for at dette skal bli så godt som mulig. Jeg håper leseren vil ha nytte av den formen jeg har gitt de utvalgte teorier, metoder og diskusjoner.

I ovenstående gjennomgang av «Eight «Big-Tent» Criteria for Excellent Qualitative Research» har ulike kvalitative sider av forskningen blitt belyst. I tillegg har validitet blitt belyst gjennom anvendelse av de syv stadiene. Studiens validitet er begrunnet med at forskningsfeltet har en stor interesse i dag. Den innsamlede empiri vil bidra til å kunne belyse et felt innen vurdering som er lite berørt. Andre vinklinger med andre metoder vil igjen kunne øke omfanget av empiri som gjør at man kan se forskningsfeltet med andre blikk. Basert på forskningspørsmålene og valgt gjennomføring av metode og anvendt empiri vil denne studiens validitet ses på som bra. Men utfordringer har blitt diskutert underveis i kapitlet. Jeg vil komme innom validitet i refleksjonen om metoder i kapittel 7.2, blant annet med et ytterligere søkelys på min forskerrolle.

## 5. Presentasjon og diskusjon av resultater

Presentasjon og diskusjon av resultatene er delt inn i fem temaer med tilhørende kategorier (kapittel 5.1 -5.5, beskrevet i kapittel 4.7, figur 31). Rekkefølgen i presentasjonen har som formål å skape en helhet og sammenheng mellom de enkelte tema. Under hvert tema presenteres resultatene i den enkelte kategori, med en påfølgende diskusjon av resultatene opp mot teori om helhetlig vurdering og tidligere forskning (presentert i kapittel 3). Bakgrunn for å gjøre diskusjoner allerede i dette kapittelet er ønsket om å ha en nærhet mellom empirien og teori og tidligere forskning. Avslutningsvis under hvert tema kommer en oppsummering av funn og som utgjør grunnlaget for den sammenfattende diskusjon i kapittel 6.

Studien er av eksplorerende karakter (redegjort for i kapittel 4.2.2 HELVYRD-designet). Det vil gjenspeiles i mitt valg om å gå bredt ut i presentasjoner og diskusjon av resultater i dette kapittelet.

### 5.1 Innhold

Tema om innhold omhandler *vurderernes faglige fokus*. Utvalgte innholdskomponenter presenteres basert på den tilgjengelige empirien. Målet er ikke å komme med en fullstendig oversikt over fagområdet, men å få fram spennvidden i det faglige innholdet som gjøres til gjenstand for vurdering. Dette er et vesentlig aspekt i gjennomføringen av en helhetlig vurdering.

#### 5.1.1 Vurderernes fokus

Figur 32 viser en oversikt over utvalgte begreper knyttet til faglig innhold som brukes i denne studien. Faglig innhold brukes i denne sammenheng noe utvidet, for å favne flere sider av en elektrikers kompetanse. De to kolonnene til venstre viser felles innholdskomponenter for alle tre vurderingssystemene, mens høyre kolonne viser komponenter som er spesifikke for fagprøven.

Felles innholdskomponenter for eksamen, fagprøve og MECVET		Fagprøve
El-Sikkerhet	Elektrofaglige	Estetikk
Dokumentasjon/underlag	Sluttkontrollskjema	Feilsøking
Utførelse og funksjonalitet	Leveres til kunde	Gjennomføring
Ulike løsninger	Selvstendighet	

Figur 32. Faglige innholdskomponenter omtalt i studien

Når det gjelder faglig innhold som del av helhetlig vurdering viser blant annet rater M3 til ulike oppfatninger. Det kan være utfordrende å se en helhet og sammenheng i forbindelse med den presenterte løsningen. Sitatet til rater M3 forteller noe om hvilken diskusjon som kan pågå.

Jeg synes foreksempel kable og vern og strømføringsevne er utrolig viktig fordi det er 95% av hverdagen til en elektriker. Mens da andre synes at funksjon er dritviktig. Jeg synes det blir helt feil og bli målt på funksjon for det kan være et uhell, men virkelig forståelse av utøvelsen av yrket vil være kabel og vern. Altså når vi sitter og diskuterer dette her så blir vi egentlig aldri enig, for da blir det jo sånn vektlegging og det blir for meg ganske vanskelig hvis han har satt en femmer på funksjon og ener på kabel og vern altså ... vi må se dette i sammenheng. (M3-P18)

Utsagnet uttrykker mangfold og kompleksitet i det faglige innholdet og åpner for en diskusjon om den faglige kompetansen i et elektrikeryrke. Nedenfor følger en presentasjon av utvalgte innholdskomponenter slik de fremkom i datasett fra alle vurderingssystemene.

### **El-sikkerhet**

El-sikkerhet er den sentrale faglige innholdskomponenten hos alle vurdererne, uavhengig av vurderingssystem. Sensor E2 uttrykker at lærlingen må være sikker for seg selv og andre først og fremst, samt vise at han/hun kan tilstrekkelig elektro og kan jobbe med andre og utføre arbeidet. Det innebærer å bruke regelverket rundt el.-sikkerhet og at det skal være sikkert for kundene. De løsningene som er valgt skal ta hensyn til utfordringer og faremomenter og de skal være praktisk gjennomførbare. Rater M1 og nemndsmedlem F3 og sensor E3 viser alle til at løsningen først og fremst må tilfredsstillе sikkerhetsforskriften, både brannsikkerhet, berøringsfare og personsikkerhet, og det må komme tydelig frem i lærlingenes arbeidsutførelse. De viser også til at sikkerhet i bruk av anlegg slik at også kunden er sikker og fornøyd. Nemndsmedlem F1 presiserer at sikkerhetsmessige feil er mye mer graverende enn andre feil knyttet til anlegget. Nemndsmedlem F2 uttrykker at «Det som må ligge i bunn er jo el-sikkerhet i forhold til at man er trygge nok til å lage et trygt nok el-sikkert anlegg og vet man har kontroll på de målingene og verifikasjonene» (F2-P16).

Sensor E2 bruker begrepet tyngdepunkt når innholdskomponenter diskuteres. Om dette uttaler sensoren at «Tyngdepunkt er et forsøk på å finne ut hva som er det viktigste for at du er på vei

til å bli en fagarbeider som kan utføre sammensatte arbeidsoppgaver på en kløktig måte» (E2-P11). Utgangspunktet er at det ikke skal oppstå en fare for arbeideren selv, og for andre arbeidere eller kunden: «... det er på en måte et el.-sikkerhetsperspektiv helt i bønn» (E2-P11). Lærlingene må legge opp et anlegg som beskytter folk mot støt og brann.

### **Dokumentasjon/underlag**

Et nøyaktig dokumentasjonsunderlag med tegninger er en vesentlig del av arbeidsutførelsen på et elektrisk anlegg for en elektriker.

En nemnd uttrykte at dokumentasjonen til en lærling på en fagprøve var utført som et lite kunstverk, og lærlingen hadde utført arbeidet på tilmålt tid og tenkt nøye igjennom de planene som ble fremlagt. Nemndsmedlem F1 uttrykte at lærlingen hadde en ekstremt ryddig og fin dokumentasjon med indeksering og at det var over forventet. Det ble begrunnet med at det var «En ganske fyldig risikovurdering og veldig mye kandidatens egne tanker om anlegget» (F1-P16). Videre ble lærlingens beregninger og gode tegninger trukket frem og dokumentasjonen var god nok til å kunne bli overlevert direkte til kunden. Det er flere elementer i forbindelse med dokumentasjoner i form av tegninger/skjema, beregninger og skriving.

### *Tegninger/Skjema*

Tegninger er et viktig arbeidsredskap og det knyttes forventninger til at de er i målestokk, har korrekte symboler, viser plassering av utstyr og kurser og at de skal kunne brukes under installasjon. Sensor E3 viser til en dokumentasjon som er selvstendig i sin utførelse med gode vurderinger og oppsummeringer og dokumenteringer gjennom skjemaer, enlinjeskjemaer, kursfortegnelser og andre tegninger som langt på vei kan benyttes. Dokumentasjonen er god nok for å anvendes i jobben, og både argumentasjon og oppbygging tilfredsstiller krav.

Nemndsmedlem F2 uttrykker sin opplevelse av tegningene til en lærling.

Tegningene er greie, oversiktlig og nummererte og de er fullt mulig å benytte dette som dokumentasjon. Klart det kunne vært tegnet penere og bedre og sånne ting, men igjen så gjør den nytteverdien sin for at man kan gå inn å se om anlegget faktisk er koblet og kan gjøre endringer og feilsøke etter skjemaene. Brukt linjal det er hyggelig. (F2-P16)

Underlaget må være utarbeidet og av en slik karakter at en annen elektriker kan bruke det i utføringen av sin jobb.

## *Beregninger*

Vurdererne i alle vurderingssystemene påpekte betydningen av beregninger og at de er en del av dokumentasjonen. Rater M4 uttrykker at *beregninger* som begrunnelser er en del av dokumentasjonen. Men Rater M4 er tydelig på at deler av dette med beregninger ikke hører hjemme hos en elektriker. Noe tilhører et høyere nivå, som for eksempel konsulenter eller saksbehandlere. Et eksempel er beregninger av belastninger og sikringsstørrelser, noe en elektriker ikke trenger å gjøre. Man kan være en elektriker uten å foreta dette, og det har ikke noe med det praktiske å gjøre. Rater M3 viser til at beregninger, som en del av det å bevise sine påstander, er vesentlig for at anlegget skal fungere, noe som kan sees på som et litt annet syn. Ulikhetene over viser at det er en meningsforskjell om ulike typer beregninger og hva som er nødvendig for å utføre en praktisk jobb.

## *Skrive*

Skriving fremstår som en sentral del i elektrikerfaget og spesielt i utarbeidelse av dokumentasjon. Sitatet fra sensor E1 uttrykker følgende om skriving og dokumentasjon:

Det var at besvarelsen er veldig fyldig og at den virker veldig reflektert og han er flink til å skrive rett og slett, og det er jo dumt at man skal måle elektrikere etter det men sånn er det jo i dag ... man må jo. (E1-P10)

Nemndene påpeker betydningen av å skrive og det er et krav i bransjen at ting synliggjøres på papir, der også lærlingen får uttrykt sin forståelse av faget. Det vurderes ikke hvordan eller hvor ryddig skriften er, men det er en fordel om dokumentasjonen er så god at den kan leveres til kunden.

## **Utførelse og funksjonalitet**

Ved eksamen vektlegges besvarelsen på en slik måte at den kan benyttes til å gjennomføre en virkelig jobb. Sensor E3 uttrykker at den må være egnet til forventet bruk. Nemndsmedlem F5 uttrykker følgende om en besvarelse ved fagprøven:

Ja, visste vel at dette var en meget kandidat hvis bare utførelsen stod i forhold til plandel og at det så ut som folk hadde gjort det. Regner med at når det ligger så mye riktig her så regner vi med at han har full kontroll og det viser seg at det hadde han. Og det er ikke alltid det er slik, men det er veldig ofte at har du kontroll på dette her så blir hele fagprøven mye enklere. Da har du satt deg ordentlig inn i det og planlegger og

kommer inn her mandag morgen og lagt opp hva du skal bli ferdig med hver dag også.  
(F5-P15)

Sitat viser at de planer lærlingen har skaper forventninger hos nemndsmedlemmene til hvordan den praktiske utførelsen blir. Ved fagprøven vil eventuelle mangler i funksjonalitet i anlegget komme tydelig frem. Noen mangler kan aksepteres men ikke det som kan medføre fare. I vurderingen av skriftlige arbeider er det vanskeligere å få innsikt i funksjonaliteten til lærlingenes løsninger, men vurdererne prøver å finne det ut ved å anvende sine praktiske erfaringer fra yrket.

### **Ulike faglige løsninger**

Et viktig bidrag kan være å se på lærlingene og deres erfaringer som en ressurs. Sensor E2 og rater M3 uttrykker betydningen av at ulike løsninger kan presenteres, og det er vesentlig som sensor å ha en åpen innstilling. Som et eksempel uttrykker rater M3 skepsis rundt løsningsromforslag knyttet til MECVET-testen (diskuteres mer i 5.3.2) siden det kan være begrensende for vurderingene. Rater M4 uttrykker at en lærling må tenke selv og at fasitsvar ikke finnes. Sensor E3 påpeker at det å vurdere besvarelser basert på åpne oppgaver vil innebære mulighet for flere løsninger. Lærlingene kan ha vært med på å bygge mange kjøkken hvor viktige punkter trekkes inn i arbeidet, og som sensor E2 uttrykker det: «... de får med seg momenter som jeg selv ikke har fått med meg.» (E2-P11). Sensor mener det er viktig å fange opp slike sider av en besvarelse og at lærlingene må få kreditt for innspillene. Sensor E2 bekrefter det ved å uttrykke at «Faget mitt består av ulike løsninger når jeg kommer inn i en fabrikk eller et rom. Det er å finne ulike løsninger og det er ikke en løsning som er riktig» (E2-P11). Samtidig viser rater M4 til en lærling som «... har flere forskjellige løsninger og gjort gode vurderinger, og har av og til flere løsninger og diskutert de litt opp imot hverandre for å komme frem til en løsning og det er bra» (M4-P18). Begge viser til åpenhet for å ivareta et mangfold av løsninger. Samtidig viser Nemndsmedlem F2 til at mennesker er forskjellige og uttrykker at «Vi har med mennesker å gjøre og vi mennesker er forskjellige og faget trenger også forskjellige mennesketyper også.» (F2-P16).

### **Elektrofaglige**

Her tar jeg utgangspunkt i et sitat fra sensor E2 for å beskrive funn rundt det elektrofaglige:

[...] de skal ta med seg kunnskap innenfor elektrofag og det kan være kunnskap i tverrsnitt på kabler, lys og varme og ulike ting inn i en praktisk situasjon, og det jeg

tror i dag vi kaller yrkeskunnskap. Er evnen til å bruke kunnskap inn i en praktisk situasjon så er mitt forhold til denne eksamen på sitt beste. Denne eleven viser evner til yrkeskunnskap da og ta med seg lyskunnskapen inn på kjøkkenet. Og det er det jeg tror jeg legger i begrepet anvende. Da skiller jeg mellom elektrokunnskap og elektroyrkeskunnskap hvor yrkeskunnskapen har med seg evnene til å bruke elektrokunnskapen inn i en situasjonsbeskrivelse. (E2-P11)

Sensor E2 viser til betydningen av å anvende kunnskapen i praktiske situasjoner og at kunnskap innebærer ulike sider av det elektrofaglige. Tre begreper brukes hvor det første er elektrokunnskap som er utgangspunktet for å anvende det inn i en arbeidssituasjon. Kunnskapen blir omtalt som yrkeskunnskap noe som fører til at man skal sitte igjen med elektroyrkeskunnskap. Dette er den kunnskapen man har anvendt inn i en praktisk arbeidssituasjon. Det er også bakgrunn for diskusjon om det er kunnskaper eller kompetanse man får vist i en slik type test som den skriftlige eksamenen er og for utviklingen av vurderingskriterier (mer om det i kapittel 5.3.2).

### **Sluttkontrollskjema**

Vurdererne legger vekt på hvordan lærlingene bruker sluttkontrollskjema ikke bare at det krysses av, men at det vises forståelse for det som velges. Sensor E1 trekker frem en lærling som gjennom sin beskrivelse viser stor forståelse av bruken av sluttkontrollskjema. Sensor sin reaksjon er at vedkommende må ha gjort det mye på jobben. Ved fagprøven uttrykker nemndsmedlem F1 følgende om bruk av sluttkontrollskjema:

Bare å krysse av her er jo verdiløst hvis du ikke forstår hva som står der. Jeg pleier å gå igjennom mye av punktene og hører om kandidaten vet hva han har krysset av på. Dette burde i utgangspunktet være lett, men veldig ofte så er forståelsen veldig dårlig i forhold til enkeltheten på skjema da, for å si det sånn. (F1-P16)

Rater M1 viser til en lærling som påpeker viktigheten av sluttkontroll. Utfordringen er at lærlingen ikke har grunnlag for å gjøre det på en skikkelig måte, siden beregningene ikke er tilstrekkelige. Forståelsen rundt det å gjennomføre sluttkontroll må være til stede.

### **Leveres til kunde**

I alle tre vurderingssystemene vektlegger vurdererne at de faglige løsningene må inneholde forslag som vil tilfredsstillende kundens behov. Rater M1 uttaler det som «Er egnet til forutsatt bruk» (M1-P17). Man skal ta hensyn til det ved planlegging av installasjonen og i

dokumentasjon. Nemndsmedlem F1 uttrykker at hovedkriteriet er å ha et anlegg som kan leveres rett til en kunde, noe som ved fagprøven ofte vil bety karakteren bestått meget godt. Rater M4 uttrykker at det er en fordel om det beskrevne kan brukes rett inn i en jobb. Forslag til elektrisk opplegg må være i samsvar med det kunden har bestilt og det må være gjennomførbart.

### **Selvstendighet**

Selvstendighet handler ifølge vurdererne om lærlingenes evne til å vise forståelse for arbeidet som skal utføres og å bruke det faglige i reelle situasjoner. Det uttrykker også at det er viktig at lærlingene reflekterer over og begrunner sine handlinger.

Sensor E2 påpeker lærlingens mangel på begrunnelser og dermed vises ikke nok kompetanse for å jobbe selvstendig som elektriker. Når det gjøres feil i anvendelse av kunnskapen i praktiske situasjoner, stiller sensor spørsmål ved om det er nok elektrokunnskaper. Samtidig uttrykker sensor at besvarelsen kan gi uttrykk for at vedkommende «[...] åpenbart har vært borti arbeidsoperasjoner enten på papiret eller gjennom praksis» (E2-P11). Sensor ser at lærlingen kunne vært med en elektriker ut på arbeidsplassen og gjort en jobb med råd og instruksjoner underveis. Spørsmålet sensor E2 stiller seg er om denne lærlingen kan være med på en elektrikerjobb, og er svaret ja, er det 2, hvis ikke, så er det stryk. Andre eksempler på lav selvstendighet er avskrift og kopi av kilder som benyttes i besvarelsene. Sensor E2 viser til en besvarelse hvor lærlingen har kopiert den tekniske løsningen fra lærebøker, en løsning som ikke vil fungere i forbindelse med den aktuelle oppgaven. Sensor E4 viser til avskrift og gjengivelse av dokumenter hvor lærlingen viser liten forståelsen for oppgaven, og sensor uttrykker at det å «Skrive rett fra forskrifter og håndbøker ikke har noe med oppgaven å gjøre i det hele tatt. Det henvises ikke til oppgaven men til generell kunnskap» (E4-P13).

Når det gjelder fagprøven viser vurdererne til selvstendighet i form av refleksjoner og forståelse ved sluttkontrollskjema og avkryssing. Som det kom frem kan avskrift og ukritisk bruk være uheldig og også bli farlig:

[...] bransjer sier generelt at det er farlig avkryssingen uten å tenke og uten å reflektere over det. [...] Men det kan se flott ut med å sette opp de kravgreiene som du kan skrive rett ut i fra en bok, uten å ha en forståelse for det du har gjort. (F2-P17)

Sluttkontroll er et eksempel på hvordan selvstendighet og dokumentasjon av egen kompetanse kan komme til uttrykk gjennom bruken av skjema.



Rater M2 viser også til samarbeid og samhandling som kompetanser knyttet til selvstendighet. Å få vurdert disse i en vurderingssituasjon kan være utfordrende, men vesentlig for yrkesutøvelsen.

Innholdselementene så langt kom til syne i alle vurderingssystemene. De tre gjenstående er mer spesifikke for fagprøven.

### **Estetikk**

Det estetiske er knyttet til den praktiske gjennomføringen av en fagprøve. Flere av nemndsmedlemmene ser på arbeidsutførelsen knyttet til oppkobling av det aktuelle anlegget. Nemndsmedlem F1 er for eksempel opptatt av å se på detaljene i utførelsen av arbeidet. Det handler både om utseende og utførelse av for eksempler koblinger. Arbeidet skal se pent ut og det skal være oversiktlig, altså estetisk og funksjonelt.

### **Feilsøking**

Nemndsmedlem F3 legger vekt på at lærlingene jobber med å finne feil i anlegget, og i prosessen bruke et skjema for å løse problemet. Nemndsmedlem F4 uttrykker at gjennom feilsøkingen får man tillit til at lærlingen forstår det som gjøres. Feil i anlegg gir noen reelle situasjoner for lærlingen, noe som gir nemndsmedlemmene mulighet til å finne ut mer om lærlingens kompetanse. Som et eksempel var det en lærling som skulle benytte utstyr i arbeidet hvor enkelte deler ikke fungerte. Lærlingen fikk ikke alt utstyr til å stemme i oppkoblingsskapet og snakket dermed med leverandøren for å få svar på problemet. Det ble sett på som bra fra nemndsmedlem F3 at lærlingen tok tak og hev seg på for å løse utfordringene gjennom å tenke selvstendig, gi begrunnelser og løse de praktiske problemene.

### **Gjennomføring**

Ved gjennomføring av fagprøven har nemndsmedlemmene anledning til å følge lærlingenes planlagte arbeide. De får innsikt i de ulike delene av den praktiske gjennomføringen, noe som gir svært viktige innspill til vurderingen av lærlingens yrkeskompetanse. Sitatet under er et eksempel på en beskrivelse fra nemndsmedlem F3 om gjennomføringen til en lærling som sliter:

Han har faktisk en god forberedelse. Så da han hadde koblet overspenningsvernet var

det koblet feil og kortslutningsskinna var ikke lagt inn. Dermed så ville ikke overspenningsvernet fungere i skapet. Så var det litt sånn estetisk faglig inni skapet som var litt rotete, [...] men det er sånn man påpeker. Så var det motoranlegget som ikke fungerte i det hele tatt og han fikk tre ekstra timer på å lete etter feil. [...] Så hadde han hovedjording ikke koblet til og det fungerte ikke. Termostat fungerte ikke heller og fikk ikke ut 230 V på skru på den. Planleggingsdelen alene gjorde at jeg trodde han skjønnte ganske mye, men når jeg så på gjennomføringen ut i fra plandel så var det som om han ikke klarte å gjennomføre det han hadde planlagt. Alle disse tingene fungerte ikke og ikke på sluttkontrollen. Kabelføringen var det knekker på som funksjonsmessig kan være dumt, og kabel så ut som om det var slengt opp og så ikke fagmessig ut. [...] sluttkontrollskjema og risikoskjema var ikke fylt ut. [...] gjorde noen målinger som han ikke fikk til og vi tok med måleapparatet opp så han kunne se litt rolig på den. Han hadde ikke forståelse hvorfor man har en kontinuitetsmåling for eksempel, altså sånn når jeg spurte enkelt om en motstand i en kabel så skjønnte han det ikke. Men magefølelsen sa at han kunne egentlig ganske mye, men var kommet feil ut og alt ble bare kaos. [...] vi måtte vurdere oppgaven og det var på måten helheten. Vi så at ting ikke ble gjennomført på de nevnte punktene, og samtalen etterpå hvor vi prøvde å dra planleggingsdelen inn i det han har gjort til dokumentasjon, og han sa jo selv at han ikke bestod. Sikkert flere punkter men dette er sånn umiddelbare. (F3-P14)

Sitatet er et eksempel på vurdering i en praktisk gjennomføring. Tilsynelatende så hadde nemndsmedlem F3 tro på lærlingen ut ifra plandel, men gjennomføringen viste mangler og utfordringen med å få det til å fungere i praksis. Under gjennomføringen kan man reelt finne ut noe om lærlingens løsninger og dermed bidra til mindre synsing rundt den faglige kompetansen. Samtidig blir det mer synlig hvor komplekst det kan være å gjøre en helhetlig vurdering av lærlingens arbeid.

#### [5.1.2 Diskusjon av vurderernes helhetlige vurdering av faglig innhold](#)

**Mangfold og kompleksitet** preger den yrkeskompetansen som skal vurderes.

Innholdskomponentene presentert i figur 33 har flere paralleller til de studiene (Petersen og Rauner: Lindberg) Rauner (2013) viser til, selv om studiene fremstår som mer generelle i sine beskrivelser. Blant annet vil metoder og teknikker være tett knyttet til vurderernes søkelys på utførelse, funksjonalitet og feilsøking. Her vil også el.-sikkerhet og elektrofaglige elementer være viktig, i tillegg kan det knyttes til valg og bruk av materialer og verktøy i

gjennomføringen. Det kan også knyttes til det Hager et al. (1994) omtaler som tekniske ferdigheter og problemløsning. Videre viser studiene (Petersen og Rauner: Lindberg) til planlegging som innebærer forståelse av oppgaven og organisering av arbeidet, noe som Hager et al. (1994) uttrykker som kunnskaper og forståelse. Dette kan ses i sammenheng med vurderernes oppmerksomhet på ulike løsninger, dokumentasjon/underlag basert på skriftlighet, beregninger og presentasjon, samt estetikk og at det kan leveres til kunde. Selvstendigheten til lærlingen berører også etiske aspekter ved yrket. De sistnevnte innholdskomponentene skiller seg ut fra de rent faglige og vil være mer knyttet til generiske ferdigheter. Forskningen til Deichman-Sørensen et al. (2011) peker også på at prøvenemndene i elektrikerfaget vektlegger el.-sikkerhet når de vurderer fagprøver. Samtidig påpeker Nore og Lahn (2015) at det er en variasjon i hva som vektlegges i vurdering av fagprøver og stiller spørsmål om hvordan flere innholdskomponenter kan trekkes inn, også mer generiske ferdigheter. I min studie kom det frem at innholdet også vil være avhengig av hva den enkelte lærling har jobbet med og lært. Ulike faglige løsninger kan bli presentert, noe vurdererne ser på som vesentlig å ta hensyn til. Forskningen til Tsagalidis (2008) peker på at både praktisk og teoretisk kunnskap legges til grunn i vurderingene, slik også vurdererne i denne studien gjør. Det er en innbyrdes sammenheng mellom innholdskomponentene, men ved at el.-sikkerhet blir så viktig vil det prege all vurdering og i en helhetlig vurdering vil det kunne virke «overstyrende» i forhold til de andre innholdskomponentene. Andre innholdskomponenter veies opp imot hverandre og tilpasses både den enkelte lærling og den aktuelle besvarelsen. Samtidig trekkes vurderingene inn mot krav til en elektrikers yrkeskompetanse og får dermed en større nærhet til yrkeslivet (Wolf, 1995). Hiim (2012) peker også på viktigheten av forankringen til arbeidspraksis. I min studie har vi sett at vurderernes ser på feilsøking hvordan de bruker den elektrofaglige kompetansen i tilknytning til den reelle arbeidsoppgaven, samt om produktet kan leveres til kunde. En slik helhetlig tilnærming ser ut til å reliabiliteten i vurderingen, sikrer valideringen av læringen, og blir en kompetanse som vil vare utover i livet (Boud & Hawk, 2003).

Vurderernes evne til å se og anerkjenne lærlingenes individuelle besvarelser kan knyttes til Erauts (2004) begrep «learning trajectories» der ulike innholdskomponenter og erfaringer bygger opp unike yrkeskompetanser. Da blir det som Johnson (2008a) påpeker, nødvendig at vurdererne evner å trekke inn og kombinere ulike innholdskomponenter i vurderingen og er åpne for ulike faglige løsninger. Dermed vil det være utfordrende å etterstrebe en standard liste med faglige forhåndsdefinerte innholdskomponenter som skal vurderes, noe som blant

annet kom frem i analysen av dokumentasjon og beregninger. Vurdererne har ofte faglige diskusjoner om innholdskomponenter på et mer generelt grunnlag, men også i tilknytning til de spesifikke vurderingskontekstene. Samtidig vil kombinasjonen av vurderernes forhåndsforståelse og deres åpenhet for innspill kreve en pragmatisk tilnærming til de enkelte besvarelsene, hvor man igjennom en prosess finner «tyngdepunktet» i det faglige innholdet.

De største forskjellene i vurdering av faglig innhold er knyttet til fagprøvens prestasjonsbaserte vurderingsform. Fagprøven har et annet utgangspunkt enn MECVET-testes mål om et kognitivt handlingspotensial. De praktiske handlingene ved fagprøven gir rom for at lærlingens kunnskap kom mer til uttrykk gjennom å vise forståelse og kunne begrunne, noe E2 omtaler som elektroyrkeskunnskap. Vurderingsformenes representativitet innebærer spørsmål om hvordan deler av det faglige innholdet kan bli vurdert. Lærlingenes selvstendighet vil ikke så lett komme til uttrykk i de rent skriftlige vurderingsformene som utførelsen og gjennomføringen av en fagprøve. Black (2000) viser til faren ved å benytte eksterne tester som ikke passer til å finne ut av hva den lærende har tilegnet seg av kunnskaper eller kompetanse. Tilpasses ikke vurderingen til den lærendes kontekstuelle virkelighet vil vurderingen fremstå som lite troverdig og det reelle faglige innholdet vil være vanskelig å få vurdert.

### 5.1.3 Funn knyttet til helhetlig vurdering ved faglig innhold

Funnene viser mangfold og kompleksitet i det faglige innholdet når yrkeskompetansen til en elektriker skal vurderes. Vurderingene omfattet dessuten både fagspesifikke og generiske (fagovergripende) kompetanser. Samtidig ligger det en viss spenning mellom vurdererne i hvordan de forstår og tolker det faglige innholdet og diskuterer seg frem til enighet om de faglige innholdskomponentene. El.-sikkerhet sto uansett sentralt i all vurdering. Det gjorde også dokumentasjon som både er en fagspesifikk og fagovergripende kompetanse. Ellers viste HELVYRD-studien et behov for en kontinuerlig diskusjon omkring gjensidigheten mellom innholdskomponenter og betydningen av å bestemme det faglige «tyngdepunktet» i den til enhver tid gjeldende kontekst. Slike diskusjoner bidrar til en større bevissthet om faglig innhold og dermed grunnlaget for helhetlige vurderinger, tilpasset lærlingens besvarelser. Vurdererne presiserer betydningen av å være åpen for ulike løsninger, og lærlingene kan bli sett på som en ressurs med praktisk erfaring fra ulike arbeidsplasser.

Mulighetene for helhetlige vurderinger avhenger noe av vurderingssystemenes representativitet, se figur 33, som synliggjør like og ulike innholdskomponenter. Vurdererne etterstrebet en helhetlig tenkning i de teoretiske testene, men de praktiske innholdskomponentene ved fagprøven, som for eksempel feilsøking, ga et utvidet grunnlag for mer helhetlige vurderinger. Diskusjonen i kapittel 5.1.2 viser at det å styre og spesifisere det faglige innholdet som en konkret liste med innholdskomponenter i testene vil svekke mulighetene for helhetlige vurderinger. Selv om det vil være et faglig innhold som alltid må tas hensyn med i vurderingene, må det gis rom for ulikt faglig innhold tilpasset de kontekstene vurderingene skjer i.

Spørsmålene om vurderingssystemenes representativitet vil bli ytterligere diskutert i kommende kapitler.

## 5.2 Rammer og kontekstuelle faktorer

Dette avsnittet omfatter presentasjon og analyse av resultater knyttet til *vurderingssystemenes form og sammenheng* samt *utforming av oppgaver og tekst*.

### 5.2.1 Vurderingssystemenes form og sammenheng

Her presenteres en gjennomgang av *formen* på de ulike vurderingssystemene, og den faktiske *gjennomføringen* og *sammenhengen* mellom vurderingssystemene.

Eksamensformen isolert sett er meningsløs ifølge sensor E2. Den kan forsvares prinsipielt når det gjennomføres praktiske vurderingsformer i forkant og etterkant av denne eksamen (tverrfaglig eksamen Vg2 og fagprøven). Men sensorene uttrykker et ambivalent forhold til eksamen, siden den ikke er egnet til å vise lærlingenes evne til å utføre arbeid og yrkesoppgaver. Samtidig påpekes det at formen på eksamen har beveget seg mot å bli casebaserte oppgaver, med situasjonsbeskrivelser som et utgangspunkt for hva lærlingene skal gjøre. Sensor E3 bemerker at eksamen er utformet slik at den tilnærmer seg mer å være den praktiske arbeidshverdagen. Sensor E3 mener dette senker terskelen for lærlingene (merk: ikke kravene) og de kan i større grad presentere praktiske løsninger som ligger nærmere deres arbeidshverdag. Sensor E2 underbygger noe av synet til sensor E3 og uttaler følgende:

Han kan si meg en god del om planlegging, det er begrenset hva man kan si om gjennomføring, kan si litt om gjennomføring. Kan si litt om å bruke yrkeskunnskapen ift kjøkken. Så ser jeg at han kan sette opp en plan på hvordan han kunne tenkt seg å gjøre denne jobben da. Så kan han si noe om gjennomføring av driftssetting som det

står i læreplan, som er i realiteten sluttkontroll i anlegget og hvor du kan si noe om kunnskap. Du kan måle feks om ledninger har ødelagt isolasjon eller ikke. [...] Så kan han si meg litt om hvilken type dokumentasjon han kan legge ved til kunden tilslutt. Det er mer en bare tegning og planlegging men den mangler jo hele gjennomføringsdelen og mangler å vurdere den konkrete jobben da. Det er ikke en helhetlig jobb ... men ja. (E2-P11)

Flere innholdskomponenter blir trukket frem, men det viktige er presiseringen til slutt om at man ikke får gjennomført en helhetlig jobb. Sensor E4 bekrefter deler av dette og viser samtidig til muligheten for at kandidaten får vist ferdigheter når det gjelder el.-sikkerhet, sluttkontroll, samværserklæringer og bruk av forskrifter og normer, samt å benytte tilgjengelig dokumentasjon.

Når det gjelder formen på MECVET-testen uttrykte rater M1 at den gir rom for å vise den aktuelle kompetansen. Vedkommende viser til erfaring med mange andre eksamensformer i yrket og sier at svært få eller ingen er så åpne som MECVET-testen. Men både rater M1 og rater M3 mener det kunne vært vedlagt en liten veiledning til oppgaven, slik at dette kunne hjulpet lærlingene i større grad til å komme i gang. Rater M2 uttrykker at formen på casene er relevante i forhold til det man møter i en arbeidssituasjon, noe som sammenfaller med opplevelsen av eksamensformen. Men rater M2 er kritisk til inndelingen av MECVET-kriteriene, noe som kan føre til oppstykking og fragmentering av vurderingsarbeidet. Det understrekes at de første dimensjonene blir mest fokusert på av lærlingene, mens de resterende får mindre fokus i besvarelsene.

Når det gjelder formen på fagprøven, er flere av nemndsmedlemmene kritiske til å gjennomføre denne ved en prøvestasjon. Det kan virke effektivt, men samtidig sees det på som et samlebandsopplegg, hvor flere av nemndsmedlemmene gir uttrykk for at det ikke er seriøst nok. Nemndsmedlem F2 mener at denne formen har blitt mer vanlig enn å ha fagprøvene ute i bedriftene, og at det kan medføre en utflating av faget. Det er viktig å presisere at ingen nemndsmedlemmer kritiserte selve fagprøven som form, men at den var organisert i form av en prøvestasjon.

Det andre området omhandler *gjennomføring*, hvor både sensor E1 og E4 nevner at eksamen utgjør den teoretiske delen for lærlinger. De mener dette er bra, siden mye av jobben for en elektriker mer og mer blir dokumentasjon, hvor de må vise at de kan mestre skriving. Sensor

E3 knytter skriving til flere deler av arbeidsoppdraget som blant annet prosjektering og sluttkontroll. Elektrikere er myndighetsbelagt til å gjennomføre dokumentasjon (ref. kapittel 5.1.1 om dokumentasjon/underlag og skriving som en del av de generiske ferdigheter). Eksamensformen oppleves som forbedret og det skriftlige uttrykket blant lærlingene har steget i kvalitet. Samtidig uttrykker sensor E4 bekymring rundt skriveferdighetene til enkelte lærlinger og spesielt de fremmedspråklige. Denne sensoren mener at det kan gjøre det verre for noen når de må skrive så mye som på denne eksamenen. Blant raterene ved MECVET-testen kommer det fram at de opplever at lærlingene kan mer enn det de skriver på besvarelsene, og en av raterene sier: «Ut i fra min erfaring så ville dette her ha vært en kandidat som hvis jeg hadde hatt han muntlig så ville han bevist uendelig mye mer enn det han får ned på papiret.» (M3-P18). Ved flere tilfeller er det ikke tvil om at lærlingene kan mer enn de skriver, blant annet uttrykker rater M3 at man sitter med en følelse av at lærlingen kan bevise sine påstander og at løsningen er riktig, selv om den ikke er optimal. Et annet element ved gjennomføring som rater M3 og rater M4 trekker frem, er tidsfaktoren ved denne testformen. De prøver å sette seg selv inn i situasjonen der de skulle svart på oppgaven på 2 timer, og det sås tvil om det ville vært nok tid til å besvare oppgaven skikkelig.

Sett opp imot den tiden han har fått på selve oppgaven så tror jeg nok hvis han hadde fått mer tid på seg, og hvis han hadde fått litt mere opplysninger om at dette skal være en del av oppgaven for eksempel, så tror jeg dette hadde blitt ganske bra, det tror jeg. (M3-P18)

Tidsfaktoren må diskuteres i forhold til flere faktorer som kan spille inn blant annet om lærlingene er vant til slike vurderingsformer, og det å kunne benytte de tilgjengelige hjelpemidlene.

Når det kommer til gjennomføring av fagprøven, inngår også skriftlighet i form av planlegging og dokumentasjon av utført arbeid. I tillegg viser nemndsmedlem F2 til en annen faktor, nemlig enkelte lærlingers lave motivasjon for å gjennomføre fagprøven på prøvestasjon og ikke ute i bedrift. Et annet aspekt som nemndsmedlem F1 uttrykker behov for, var å ha mer kontroll på hva lærlingene gjorde, slik at vurderingsgrunnlaget blir basert på lærlingens eget arbeid. Det å ta med seg ting hjem for å forberede seg ga ikke mulighet for å følge opp lærlingen. Det vil jo komme frem under fagprøven om de kan det eller ikke, men det var i enkelte tilfeller oppdaget at deler av dokumentasjonen for to lærlinger kunne være identiske. Nemndsmedlem F1 uttrykte å ville ha fagprøven mer som en eksamensform hvor

man fikk eliminert muligheten for at lærlinger kopierte av hverandre. Videre ble det av et nemndsmedlem rettet kritisk blikk på at lærlingene bruker ferdige maler eller beregningsprogrammer som gjør mye av planleggingsjobben. Nemndsmedlem F1 mente dette er en svakhet ved prøveformen, men påpekte at det er en diskusjon innad i nemndene om dette er positivt eller ikke. Samtidig kom det tydelig frem at nemndsmedlemmenes intensjon var å gi en så rettferdig vurdering som mulig innenfor de rammer som gjaldt for fagprøven.

To av sensorene viser til at det som oftest er en sammenheng mellom resultatene på eksamen og på fagprøven. Sensor E2 uttaler at erfaringsmessig viser det seg at kunnskapen de uttrykker på en skriftlig prøve også kommer til uttrykk i gode handlinger i praktiske oppgaver. Men sensor E2 uttrykker at terskelen er litt lavere for å stå på eksamen enn på fagprøven. De begrunner det med at man er mest opptatt av at de har fått med seg en del kunnskap som det er sannsynlig at de kan bruke i en arbeidssituasjon (anvende kompetanse). Sensor E3 uttrykker at «Kunnskapsmessig så ligger kandidatene veldig nære hverandre når det gjelder nivå her og på fagprøven, og fagprøven kommer 6 mnd etterpå» (E3-P12). Samtidig påpeker sensor E3 at det motsatte også kan være tilfelle. Lærlingen imponerer med gode praktiske ferdigheter uten at det kom til uttrykk på eksamen. Sensor E4 uttrykker at når dokumentasjonskravet til en elektriker har blitt høyere kan en slik type eksamen forsvares, og den vil fremstå som utfyllende sammen med en fagprøve. Rater M3 kan se likheter mellom MECVET-testen og fagprøven og spesielt i måten man tenker vurdering. Rater M2, derimot, påpeker forskjeller i vurderingen i forhold til fagprøven der grunnlaget for vurdering er mer omfattende. De to raterene har begge erfaring med fagprøven og ser noe ulikt på disse to systemene. Her er det snakk om graden av likhet og ulikhet og viljen til å kunne se sammenhenger som kommer frem blant disse to raterene. Det kommer frem synspunkter om sammenhenger mellom vurderingssystemene, samtidig som det påpekes forskjeller i hva man kan få vist gjennom de ulike former for vurdering. Det viser behovet for å ha flere og ulike vurderingsformer for å fange opp mangfold og kompleksitet i yrkeskompetanse.

### 5.2.2 Utforming av oppgaver og tekst

Her har jeg sett på *utforming* av oppgavene og graden av *åpenhet* på de anvendte oppgavene. Sensorene E2 og E3 uttrykker at oppgavenes *utforming* som en arbeidsordre eller et arbeidsoppdrag ligger nær den praktiske arbeidshverdagen. Det er en måte å beskrive en jobb på som er lett å kjenne igjen. Sensor E2 uttrykker at disse oppgavene er noe av det bedre som er utarbeidet, siden de består av en kortfattet stikkordsmessig arbeidsordre. Tidligere hadde oppgavene mye tekst som skulle beskrive alt så grundig som mulig, slik at det skulle gi en



opplevelse av å stå der ute på anlegget. Dette opplevde lærlingene som fullstendig meningsløst, fordi du får ingenting ut av teksten og ser ikke situasjonen. Sensor E2 uttrykker dilemmaer knyttet til lærlingenes yrkeskunnskap presentert i besvarelsene, om yrkeskunnskapen kan være vanskeligere å kjenne igjen på grunn av oppgavenes formuleringer. Noe av kritikken mot oppgavene dreier seg om mangel på informasjon i oppgavetekstene. Både tekst og informasjon må være mer tydelig. Sensor E2 sier følgende: «Og av og til er oppgaven mer uklar enn jeg har forutsatt at de er da. Av og til ser jeg hva hensikten med hva oppgaven har vært, men så ser ikke kandidaten hensikten med oppgaven». (E2-P11). Enkelte sensorer kommenterer at bredden i oppgaven blir for snever og favner bare deler av fagfeltet. Den dekker ikke hele læreplanen på Vg3-nivå, noe som peker tilbake på diskusjonen om faglig innhold.

Ved fagprøven er det også utfordringer med oppgavene. Nemndsmedlem F3 opplever dem som rotete og kunstige sett i forhold til hvordan prøvebåsene på prøvestasjonen er bygd opp. Nemndsmedlem F4 etterlyser mulighet for nemndene til å medvirke i utvikling av oppgavene for å få mer eierskap til prøvene. For å sikre kvaliteten på fagprøven er nemndsmedlemmenes spørsmål til kandidatene vesentlige som del av vurderingsgrunnlaget. Nemndsmedlem F5 viser til den stadige utviklingen i elektrikerfaget, og dermed behov for videreutvikling av oppgavene. Den opplevde relevansen til dagens oppgaver oppleves som både og. Nemndsmedlem F2 er kritisk til prosessen og mener at oppgaven er lagd etter minste felles multiplum. Det har vært en lang diskusjon mellom nemndene og bedriftene, og det har endt opp i noe som per i dag ikke sees på som en god løsning. Nemndsmedlem F1 påpeker at oppgavesettet som brukes kunne både vært byttet ut og rotert oftere.

Ved MECVET-testen uttalte rater M3 at det er mye tekst som kanskje ikke er nødvendig å ha med, og som eventuelt vil virke forstyrrende for lærlingen i arbeidet med å løse oppgaven. Samtidig mener rater M3 at det mangler en god del opplysninger i oppgaveteksten for at den kan løses fullt ut. Selve oppgaveteksten bør være ganske direkte, noe som hadde hjulpet lærlingene. Rater M1 mener oppgavene er realistiske og poengterer viktigheten av å trekke sammenligninger til det å utføre et oppdrag i virkeligheten. Oppgavene sees som relevante men lærlingene har problemer med å se alle aspektene med jobben når de får en så åpen case-beskrivelse. Rater M1 mener beskrivelsen av oppdraget ville vært annerledes i virkeligheten. Rater M1 sine betraktninger om oppgaver handler også om hva lærlingene er vant med, og er noe en vurderer må ta hensyn til i vurderingsarbeidet.

Når det gjelder oppgavenes *åpenhet* uttrykker sensorene E3 og E4 at de er veldig åpne og sensor E4 sier at sentrale temaer dermed kan bli oversett. E3 sier at åpne oppgaver kan bli omfattende for en lærling som sitter i en presset eksamenssituasjon og skal tenke på hvordan det skal planlegges, gjennomføres og dokumenteres. Sensor E2 sier at de får et større karakterspenn med disse oppgavetyperne og det kan være med på skille lærlingene i større grad. Det kan skyldes at det ikke er noe fasit svar og det er nærmere en praktisk hverdag, hvor de får frem lærlingenes reelle kompetanse.

I mitt materiale kom det ikke til uttrykk noe spesifikt omkring åpenhet ved oppgavene på fagprøven. Når det gjelder MECVET uttrykker rater M4 en bekymring for at åpne oppgaver fort kan føre til at lærlingene ikke kommer i gang med å løse den. Det kan også innebære at enkelte av dimensjonene, som sikkerhet og miljø, blir dårlig besvart. Rater M4 uttrykker en mulig årsak til dette:

Du ser de fleste faller igjennom på miljøkrav og hms sånne ting. Du gjør det bare og på jobben drar du på deg hjelmen og verneskoa og det er ikke noe du gjør for at det skal stå skrevet, det er helt automatikk på dette. Og det tenker du ikke på når du skal skrive en oppgave. (M4-P18)

Rater M3 trekker sammenligninger til Vg3-eksamen for lærlinger og at det er underforstått at beregninger og slikt skal være med, samt miljø og arbeidsprosesser. Noen kritiske synspunkter knyttes til at oppgavene er på et for høyt nivå, eller ikke relevant. Man trekker linjer til prosjektering og planlegging og det uttrykkes at «De får ikke gjøre en slik jobb alene sånn da er det grensen mellom prosjektering og planlegging.» (M2-P17). Vurdererne påpeker at åpenhet i oppgavene både gir muligheter og utfordringer og ønsker at de som utformer oppgavene tar hensyn til dette.

### 5.2.3 Diskusjon av rammer og kontekster for helhetlig vurdering

Både *vurderingssystemenes form og sammenheng*, og *utformingen av oppgaver og tekst* har innvirkning på vurderernes muligheter for helhetlige vurderinger.

Vurderingssystemens form kan knyttes til Biemans et al. (2009), som uttrykker at poenget med å ha kompetanse-basert utdanning vil være å skape mindre gap mellom skole og arbeidsmarked. Samtidig presiserer Wolf (1995) at kompetanse-basert vurdering legger mer vekt på vurdering i ikke-akademiske sammenhenger, noe som betyr å være mer nært

yrkeslivet. Engh et al. (2007) peker på gyldigheten til hva en vurdering kan avdekke som vesentlig. Noen former kan avdekke alt for lite av den kompetansen som skal vurderes. Resultatene viser at de mer åpne casebaserte formene til vurderingssystemene blir sett på som noe positivt, med en intensjon om å være nærmere yrkeslivet. Samtidig kom det frem et kritisk blikk på i oppdelingen av dimensjonene på MECVET-testen, noe som var utfordrende for lærlingene å forholde seg til. Bruk av prøvestasjon ved fagprøven ble ikke sett på som positivt, og det stilles spørsmål om relevans. Wolf (1995) og Hiim og Hippe (2001) viser til betydningen av at vurdering gjøres i en setting nærmest mulig den aktuelle arbeidssituasjonen. Ecclestone (2010) peker på at kvaliteten på vurderingen må baseres på å bruke det beste fra arbeidslivet og skole. Kritikken rundt bruk av prøvestasjon ved fagprøven kan nyttes til det Wolf (1995) skriver om viktigheten av å tilpasse vurderingsformene til den reelle situasjonen, ellers kan vurderingene bli sett på som uvesentlige. Vurderingssystemene må være «suited to the task» (Dobson, 2010). Olsen (2009) peker på nytten og funksjonen til vurderingssystemene og den intenderte hensikt. Vurderingen skal ikke bare handle om læringsresultater, men også om kompetanse som kan anvendes (Dobson, 2010). Baartman et al. (2007) sitt kriterium om *autensitet* peker på at vurderingssituasjonen må reflektere det som skjer i den reelle arbeidssituasjon. Black (2000) støtter det, og sier samtidig at det vil kunne styrke sammenhengen mellom den formative og summative vurderingen, og opplæringen kan få en sterkere tilknytning til den reelle arbeidssituasjonen og den aktuelle kompetansen (Tanggaard, 2004). Det kan skape en økt forutsigbarhet i vurderingsformene (Tanggaard & Elmholdt, 2008; Eraut, 2004). Samtidig viser Friche (2010, 2013) til betydningen av de sosiale og kontekstuelle faktorenes påvirkning på vurdering. Hun viser til at den prosessorienterte vurderingen uteble, noe som begrenset den helhetlige vurderingen. Dermed kan det hemme sammenhengen mellom den formative og summative vurderingen, og igjen nærheten til yrkeslivet. Casebaserte oppgaver som prøveform etterstreber å være autentisk, men som nevnt ble det av vurdererne i varierende grad oppfattet som godt nok i alle systemene, og spesielt ikke i fagprøven. Man kan stille spørsmål ved hvor autentisk en prøvestasjon kan utformes og hvor relevant og meningsfylt det vil fungere for lærlingen. Samtidig må spørsmålet stilles om hvor autentisk casebaserte vurderingsformer kan bli. Fagprøven kan knyttes til prestasjonsbaserte former (Hayton & Wagner, 1997) med fokus på det praktiske i gjennomføringen. De to andre kan sees på som casebaserte former (Segers et al., 2008) og som Lounsbury og Pittenger (2011) relaterer til den konstruktivistiske tenkningen med fokus på den lærendes aktive rolle både i oppgaveløsning og vurdering. Betydningen av å ha forskjellige vurderingssystemer støttes av flere forskere (Gipps, 1999;

Black, 2000; Shepard, 2000). Å ha ulike systemer for vurdering kan bidra i prosessen med å få frem flere sider av en kompetanse. De utvalgte vurderingssystemene har ulike utgangspunkt og er sammensatt av ulike deler og preger dermed den helhetlige tenkningen rundt vurdering. I den sammenheng ble det uttrykt at eksamen ikke kunne stå alene for å vurdere kompetansen. Vurdererne i studien uttrykker behov for å ha flere vurderingsformer. Både Hayton og Wagner (1997) og Hager et al. (1994) uttrykker at penn og papir tester ikke er tilstrekkelige for å vurdere en kompleks kompetanse, mens Rauner (2013) i større grad forsvarer det gjennom å peke på at man tester på et mer konseptuelt nivå. Blant vurdererne på både eksamen og MECVET kom det frem oppfatninger om hva man kunne få vist ved de ulike vurderingsformene, samtidig som de presiserte begrensningene. Sensor E2 uttalte at man ikke fikk vurdert en helhetlig jobb. Transparens (Baartman et al., 2007) og gjennomsiktighet (Engh et al., 2007) bidrar til å bevisstgjøre hva som kan bli vurdert. Engelsen (1998) viste til behovet for åpenhet rundt fagprøven i det offentlige rom. Man må vite hva prøvene skal teste ut. Åpenhet og transparens kan også bli utfordret siden vurdering av en kompleks yrkeskompetanse også kan gjøre vurderingen mindre transparent. Engelsens fremstilling knyttes til kriteriediskusjonen i kapittel 5.3. Et annet aspekt ved transparens er læringsperspektivet ved vurderingsformer. Spørsmålet er om de lærende kan bedømme seg selv og sin egen læring i en gitt vurderingssituasjon (Baartman, 2007), noe som kan sees på som en styrke i utformingen av vurderingene. Avslutningsvis vil jeg nevne at Eraut (2004) påpeker umuligheten i å utvikle det perfekte vurderingssystem, spesielt ved vurdering av en kompleks yrkeskompetanse.

*Gjennomføring* av vurdering, har en styrke når det skjer nærmest mulig en reell handling jfr Baartman et al. (2007) sitt kriterium om *direkthet*. I det lyset kan man tolke det slik at fagprøven er mer direkte enn de to andre vurderingssystemene, som kan oppleves mer indirekte. Betydningen av skriftlighet ved de ulike formene for vurdering framkom som sentralt (beskrevet i kapitell 5.1.1 også). Skrivning er vesentlig for lærlingene, men formen for skrivning kan se ut til å ha betydning. Å skrive for å utvikle en dokumentasjon kan sees på som noe annet enn det å skrive for å forklare hvordan man vil løse oppgaven. Raternes opplevelse ved MECVET-testen viste at de følte at lærlingene hadde mer kunnskap enn det som sto skrevet, og muntlige høringer kunne gitt mer utfyllende svar. Utvikling av dokumentasjon kan virke mer direkte og være nærmere en praktisk yrkeshverdag enn det å be om redegjørelser for hvordan oppgaven skal løses. To andre faktorer ved gjennomføringen kan knyttes til Baartman et al. (2007) sitt kriterium om *rettferdighet*. Tidsfaktoren ble sett på som kritisk ved

MECVET-testen. Det andre var lærlingenes behov for tilgang til hjelp og støtte og dette var ulikt ved vurderingssystemene. Ved fagprøven kom det frem noe kritikk angående organisering av forberedelsene og bruk av hjelpemidler. Både Engh et al. (2007) og Hiim og Hippe (2001; 2009) fremhever betydningen av at lærlingene får vist hva de kan. Hiim trekker frem behovet for å differensiere vurderinger for å tilpasse det som skal vurderes både til tidsperspektivet og de mulighetene lærlingene har for hjelp og støtte. Rauner (2013) og Hayton og Wagner (1997) påpeker viktigheten av å lage oppgaver som kan tilpasses ulike nivåer, noe som kan ivareta differensieringen. Selv om man har åpne vurderingsformer som et utgangspunkt vil det ikke automatisk sikre en rettferdighet i vurderingen.

Vurderingsformen må baseres på kompetansens natur ifølge Wolf (1995) og det kan være behov både for flere vurderingsformer og for å se *sammenhenger* mellom resultatene på tvers av vurderingssystemer. Bruk av flere vurderingssystemer vil gi et større mangfold, noe flere forfattere forfekter (Baartman et al., 2007; Gipps, 1999; Shepard (2000); Dobson, 2010). De ulike vurderingsformene kan frembringe ulike sider av en kompetanse og utfylle hverandre. Baartman et al. (2007) sitt kriterium om *reproducibility of decisions* og prinsippet om multiple vurderingsformer kunne være med på å stadfeste kompetansen en lærling kan ha (ref til Engh et al. (2007) om gyldighet). Igjen gir det grunnlag for kriteriet om *sammenlignbarhet* hvor resultater på eksamen og fagprøve vil være interessant for å sette vurderingssystemene opp imot hverandre, noe Lai (2011) mener kan være et nyttig element. Likheten i resultater mellom eksamen og fagprøve er ikke nødvendigvis uttrykk for at de fanger opp de samme sidene ved en yrkeskompetanse. Eksempelvis vurderte de i større grad de mer generiske feridighetene ved fagprøven. Et annet aspekt var at terskelen for å få bestått eksamen var lavere enn ved fagprøven. Det enkelte vurderingssystem fanger opp ulike sider ved kompetanse, noe jeg vil komme nærmere inn på i de kommende temaer. Et siste element når det gjelder sammenheng er at vurderingen gjøres av mennesker, noe Hayton og Wagner (1997) understreker betydningen av ved kompetansevurdering. Rollen som vurderer i de tre ulike systemene preges av hvordan vurdereren forholder seg til det som skal vurderes og hvilke rammer som er gitt for vurderingsarbeidet. I denne studien hadde vurdererne enten direkte eller indirekte (gjennom besvarelsene) kontakt med lærlingene. Det har også betydning for å se på sammenhengen mellom vurderingssystemene og resultatene av vurderingene Det vil bli ytterligere diskutert under forutsetninger i kapittel 5.5.

Når det gjelder *utforming av oppgaver og tekst* peker Lai (2011) på at prestasjonsbaserte oppgaver skal kunne måle de evnene de lærende har på en dypere måte. De praktiske

ferdighetene og den helhetlige forståelsen til lærlingen må bli vurdert. I alle oppgaveutformingene etterstrebes det en *autentisitet* for å utfordre lærlingene i størst mulig grad med det som møter dem i yrkeslivet noe både Wolf (1995) og Rauner (2013) kaller å ha det «decouplet» fra skoler. Hayton og Wagner (1997) viser til betydningen av å ha vurderinger som er relevante, hvor prestasjonsbaserte oppgaver med praksis er helt sentralt for å kunne gi en god nok vurdering. Fagprøven legger opp til det, men selve oppgaveutformingene fikk i større grad kritikk fra prøvenemndene. Forskningen til Lahn og Nore (2015) viste at utviklingen av oppgaver i nemndene til fagprøven var realistiske, men variasjonen var stor, alt fra åpne komplekse til løsrevne deloppgaver. Betydningen av om fagprøven gjennomføres på prøvestasjon eller i bedrift vil også kunne påvirke. Segers et al. (2008) og Rauner (2014) viser til bruk av casebaserte oppgaver som grunnlag for kompetanse- og helhetsorientert vurdering. Oppgavene til den skriftlige eksamen og MECVET-testen er varianter av dette. Analysen viser forskjeller i mengde tekst der vurdererne påpeker at for mye tekst kan virke forvirrende for lærlingen. De påpeker også ulikhetene mellom eksamen og MECVET der eksamen har mer direkte og kortfattet i arbeidsordren. Samtidig viser Jacobs (2016) til at oppgavene benyttet i KOMET-modellen har en høy score på relevans i forhold til eget yrke. Spørsmålet som kan stilles i den sammenheng er hvor vant elevene i tysk skole er med den oppgaveformen. Å kjenne til formen på oppgaver vil være viktig for at den lærende selv kan være aktiv og utvikle sin forståelse (Lounsbery og Pittenger, 2011). De casebaserte oppgavene legger opp til det, og kanskje spesielt når de har form av en arbeidsordre. Boud og Hawk (2003) påpeker at oppgavene ikke bare skal måle øyeblikket, men også ivareta et fremtidsperspektiv for læring. Caseoppgaver fordrer en mer aktiv deltakelse fra lærlingene, som kan gi dypere læring og forståelse hos lærlingene. Utforming av oppgaver har også betydning for hvordan yrkeskompetanse kan avdekkes og vurderes (Hiim og Hippe, 2001; 2009). Vurdererne i studien stilte spørsmål om de åpne oppgavene avdekket nok i forhold til kompetansemål og dimensjoner. Wolf (1995) peker på at vurdering vil foregå på bruddstykker av en kompetanse, og at det vil være utfordrende å få dekket alt i en vurdering.

*Åpenhet* i oppgavene kan sikre nærhet til reell yrkesvirksomheten (Hiim og Hippe, 2001) og at lærlingene får vist hva de kan (Hayton & Wagner, 1997; Rauner, 2013). Slik åpenhet er mest etterstrebet i eksamen og MECVET-testen. En større del av karakterskalaen ble benyttet ved bruk av arbeidsordre som oppgaveform ved eksamen, og dermed ble det en annen måte å skille de beste fra de mindre gode. Samtidig kom det frem på både eksamen og MECVET-

oppgavene at en slik åpenhet krever både nok og riktig informasjon og at dette til dels kunne mangle. Noen av vurdererne opplevde det også problematisk å forstå selve oppgaveteksten når det manglet informasjon. Slike uklarheter kan påvirke *transparensen*, slik Baartman et al. (2007) trekker frem, men vurderingssystemene må allikevel kunne tilpasse seg virkeligheten i yrkeslivet og forårsaker problemer og svekke tilliten (Wolf 1995). Lærlingene må oppleve det som *meningsfylt*, og det må ha en verdi. Segers et al. (2008) viser til en dypere læring ved bruk av casebaserte oppgaver i vurderingen når de lærende forstår kravene, samtidig som *rettferdighet* skal ligge til grunn (Baartman et al., 2007). Utviklingen av vurderingssystemer er problematisk og komplekst og ifølge Struyven et al. (2006) bør det være en triangulering av metoder. Vurdererne i studien viste til at åpne oppgaver kunne bli for omfattende for lærlingene å forholde seg til og at nivået på krav til kompetanse for å løse oppgavene kunne virke for høyt. Samtidig var de positive til de mulighetene åpne oppgaver ga for at lærlinger kunne utvikle nye og ulike løsninger som vurdererne måtte forholde seg til og som gjorde vurderingsarbeidet mer utfordrende. Synspunktene fra vurdererne er til en viss grad motstridende, men samtidig handler det om en balansegang hvor lærlingene blant annet må bli vant med vurderingsformer av et slikt format. Så om utformingen med helt åpne oppgaver er svaret ved kompetansebasert vurdering, eller om mer tekst med informasjon og delvis instruerende spørsmål kan gi vel så gode vurderingssituasjoner, er det ikke noe enkelt svar på. Men det må etterstrebnes en utforming av vurderingssystemer som lærlingene kjenner seg igjen i og som kan bety ulike varianter og en tilpasset tilnærming. Studien viste høyst ulike former og ulik grad av åpenhet i de anvendte oppgavene.

#### 5.2.4 Funn knyttet til helhetlig vurdering ved ulike rammer og kontekstuelle faktorer

Alle tre vurderingssystemene har en tydelig intensjon om å være nær yrkeslivet. Samtidig fremhevet vurdererne de casebaserte formene med mulighet for mer åpne vurderinger, men at det også var begrensninger ved eksamen og MECVET-testen når det gjaldt grad av representativitet i forhold til yrkeslivet. En skriftlig eksamen kan ikke stå alene som en vurderingsform av yrkeskompetanse. Fagprøvens prestasjonsbaserte form avdekker flere sider av en yrkeskompetanse, men bruken av prøvestasjon ble kritisert for ikke å være nært nok virekelig yrkesutøvelse. Større åpenheten i vurderingsformene minsker graden av transparens og forutsigbarhet, og øker kompleksitet og variasjonene. På den måten utfordres vurderernes vurderingsarbeid samtidig som det gir større rom for helhetlige vurderinger.

Alle tre vurderingssystemene inneholdt krav til skiftlighet med direkte relevans for yrkeslivet, både i form av dokumentasjon og faglige redegjørelser. Skrivning av dokumentasjon har en mer direktehet mot yrkeslivet. Skrivning om faglige redegjørelser på eksamen mente sensorne at hevet seg. Men en muntlig del i tillegg hadde bidratt positivt for flere lærlinger.

Tidsfaktoren ble opplevd som kritisk ved MECVET-testen. I tillegg var det en diskusjon om anvendelse og tilgang på hjelpemidler ved alle tre vurderingssystemene.

Noen av de som vurderte i flere systemer pekte samtidig på at de så en sammenheng mellom resultatene på en skriftlig eksamen og den mer praktiske fagprøven. Sammenhengen kan uttrykke noe om vurderingssystemenes representativitet. Vurderere så fordelene ved å ha multiple vurderingssystemer med komplementære funksjoner i vurderingen av en yrkeskompetanse for å sikre helhetlig vurdering, noe tidligere forskning støtter.

Oppgavene og tekstene var i stor grad autentiske og aktiviserte lærlingene til å tenke hvordan de ville ha utført arbeidet i det virkelige yrkeslivet. Eksamensoppgavene kom best ut med oppgaver utformet som en arbeidsordre, noe som også var gjenkjennelig for lærlingene. MECVET-testen hadde for mye informasjon i oppgaveteksten og kunne forvirre lærlinger som ikke hadde trent på oppgaveformatet, mens oppgavene ved fagprøven kom dårligst ut da de i større grad var mer styrende og for konkrete. Det underbygges med at nemndsmedlemmene etterlyste en større grad av medvirkning for å følge fagets utvikling. Vurdererens muligheter for helhetlig vurdering var avhengig av oppgaveformatet og lærlingenes muligheter for å vise hva de kunne og foreslå løsninger basert på erfaringer fra yrkeslivet. Denne studien viser en utvikling mot bruk av åpne oppgaver og dermed muligheter for individuelle løsninger og stiller nye krav til vurdererens vurderinger. Oppgaveteksten og formatet kan ikke alene være avgjørende for vurderernes muligheter for helhetlige vurderinger.

Utforming og åpenhet i oppgaver og tekster, slik det foreligger i denne studien, vil i lys av Baartman (2007) sine kriterier sies å underbygge for helhetlige vurderinger. Oppgavene benyttet i eksamen og MECVET har kommet lengst i retning av å være casebasert og åpne.

### 5.3 Mål og kriterier

Dette delkapitlet handler om vurderernes bruk av *kompetanse og faglige redegjørelser* i vurderingsarbeidet. Det handler også om utvikling og bruk av *vurderingskriterier*. Avslutningsvis omtales *karakterens betydning* i vurderingsarbeidet. Formålet med



vurderingssystemene (beskrevet i kapittel 4.1) er å få avdekket kompetanse. I foregående delkapittel ble behovet for å ha ulike vurderingssystemer belyst. Vurderingssystemene har også innvirkning på hvilke mål og kriterier som kan anvendes i vurderingen.

### 5.3.1 Kompetanse og faglige redegjørelser

I vurderingsarbeidet skal vurdere forholde seg til fastsatte standarder i form av kompetansemål eller kriterier. I analysen av vurderernes praksis har jeg i denne sammenheng sett etter hvordan de har *forholdt seg til* kompetansemål og MECVET-kriterier, samt deres faglige redegjørelser for kompetanse på *ulike nivåer* i første rekke som fagarbeider og dernest grenseoppganger til andre nivåer som saksbehandler/teknikernivå eller ingeniørnivå.

Sensor E1 på eksamen uttrykker at kompetansemålene ikke benyttes direkte inn i vurderingsarbeidet, men etter mange års erfaring har man «snerten av de inne, men kan de ikke på rams». Sensor E1 vurderer aldri besvarelsene direkte opp imot kompetansemålene, noe man ikke ser vitsen med da de oppleves for vide og åpne og lite spesifikke. Sensor E3 forholder seg til læreplanen i forhold til lærlinger på arbeidsplassen og mener at den sånn sett benyttes. Vurderer E3 uttaler at «[...] jeg sitter ikke med læreplan foran meg å leser kulepunkt målene samtidig som jeg sensurer. Men jeg har den i bakhodet og vi jobber jo med den» (E3-P12). Sensor E2 viser til at gjennom mange diskusjoner opp igjennom årene uttrykkes det:

[...] at jeg kan min læreplan og den ligger i ryggmargen så skyldes det at vi har diskutert den i så mange år, og at jeg opplever at vi har en stor grad av enighet om hva som ligger i å planlegge, gjennomføre, driftsette og dokumentere en jobb om den er på lys eller om den er på varme eller om den er på et sikringsskap». (E2-P11)

Opplevelsen av felles forståelse av kompetansemålene underbygges av de andre sensorene. Det kommer fram at kompetansemål mest ser ut til å utgjøre en ramme for vurderingene. Tilvarende uttalelser kom også fra nemndsmedlemmene på fagprøven. Men både nemndsmedlem F1 og F2 opplevde at kompetansemålene kunne virke å være på et høyt nivå. F2 uttrykker at «Vi kunne laget fagprøver her etter kompetansemålene i Vg3 plan og med hevet hodet sagt dette skal dere kunne. Og vi hadde hatt tolv strykkandidater her hver gang ... kompetansemålene er høytsvevende og jeg synes de ligger veldig høyt.» (F2-P16). Samtidig trekker sensor E2 frem opplevelsen av en todelt læreplan. Noen av kompetansemålene oppleves som en reell kompetanse mens andre kompetansemål ses mer på som rene

kunnskapsmål. Det kan ha betydning for grunnlaget ved gjennomføringen av vurderingen.

Raterene i MECVET-testen viser til at de ser en sammenheng mellom kompetansemålene og MECVET-kriteriene. Rater M3 uttrykker om sammenhengen at «Ja, det tror jeg nok helt klart. Jeg ville nok ha kunne koble alle disse punktene her opp imot kompetansemålene ja, det tror jeg faktisk ville ha klart.» (M3-P18). Det kommer også til uttrykk blant raterene at MECVET-kriteriene kan ligge på et for høyt nivå. Men rater M3 uttrykker om MECVET-kriteriene og kriteriespesifikasjonene at «Ingen av disse kulepunktene her som er tull ikke sant, alle er jo helt klart reelle.» (M3-P18). Det kan virke motstridende mot hva ratere uttalte om et for høyt nivå, men kan forstås dithen at innholdet og betydningen av MECVET-kriteriene er forståelige. Samtidig oppleves noen av dem som for vanskelige og eventuelt irrelevante for en lærling. Spesielt gjelder det kriteriene knyttet til organisering av arbeid, miljøansvar og effektivitet og lønnsomhet. Uttalelsene fra raterne samsvarer med opplevelsene knyttet til kompetansemålene og hvordan sensorene og nemndsmedlemmene forholder seg til dem i vurderingen.

Når det gjelder vurderernes egne faglige redegjørelser og *ulike nivåer* viste sensor E3 til at en fagarbeider kan utøve yrket med et lavere kunnskapsnivå enn det som ble vist på noen besvarelser. Dette ble begrunnet med at fagarbeideren utøver faget praktisk og tar praktiske avgjørelser i det praktiske arbeidet. Samtidig ble det presisert at praktisk yrkesutøvelse bygger på en plattform med teori i bunn. Ved enkelte anledninger ble det uttrykt av sensorene at når lærlinger ble vurdert til toppkarakterer, ble det ofte argumentert med at de skulle videre i utdanningsløpet og bli noe mer enn «bare» fagarbeider. Sensor E3 viste til en besvarelse med toppkarakter og sammenlignet dette med et teknisk fagskole-nivå, kanskje et ingeniørnivå. Men sensor E3 hadde sett at gode teoriprøver ikke alltid ble omsatt i gode praktiske ferdigheter og uttaler at «[...] det er veldig vanskelig å lese ut av den oppgaven her om vedkommende kan gjøre de riktige valgene i en praktisk arbeidssituasjon». (E3-P12). Dette henger også sammen med det som ble diskutert om sammenheng i kapittel 5.2.1.

Blant nemndsmedlemmene ved fagprøven ble det også uttrykt noe om grenseoppgangen mellom en fagarbeider og andre nivåer. Ved et par gode besvarelser uttalte nemndsmedlem F5 at disse lærlingene mest sannsynlig ville bli ingeniører: «Begge de to siste er jo folk som vil gå videre. Så det er folk som ser fremover og blir ikke lenge som fagarbeidere, de vil bli ingeniører.» (F5-P15). Sammenligninger med andre nivåer ble en del av vurderernes faglige redegjørelser og grunnlag for vurderinger. Rater M4 viser til en konkret besvarelse og uttaler

at dette er mer for saksbehandlere og ingeniører og ikke for fagarbeideren. Det begrunnes med vektleggingen på planlegging og beregninger som er på et annet nivå enn det som gjelder for en utøvende elektriker, og har lite med den praktiske biten å gjøre. Enkelte av beregningene kan man klare seg fint uten, og rater M4 mener disse kan knyttes til andre nivåer. I et annet tilfelle viser Rater M4 til en besvarelse hvor en lignende logikk kommer til uttrykk, men med motsatt fortegn:

Jeg legger vekt på at denne kandidaten her gjør sånn som jeg ville gjort ute. Og når jeg leser dette her så kan jeg ta med meg papiret ut og utføre denne jobben som står her. Fordi at dette er en praktisk elektriker og ikke bare en som er interessert i å regne og skal bli noen ingeniør og dette er en som har tenkt. (M4-P18)

Gjennom Rater M4 sine eksempler kan man tolke det dithen at det i det første tilfellet blir for teoretisk og overordnet og mer styrt mot selve planleggingen. I det andre tilfellet vises det mer til den konkrete utøvelsen av yrket. Dette kan være et eksempel på grenseoppgangene mellom de ulike nivåene.

### 5.3.2 Vurderingskriterier

I den norske konteksten er det krav til utvikling av kriterier som grunnlag for vurderingsarbeidet. Dette gjelder både eksamen og fagprøven. For MECVET-testen fikk vurdererne utdelt ferdige vurderingsskjemaer med kriterier fra KOMET-modellen tilpasset norske forhold. I denne delen av studien har jeg analysert hvordan vurdererne *benytter kriterier* og kriterienes *stramhet*.

Ved å *benytte kriterier* uttrykte sensor E1 at forventningene var klare til hva en lærling skulle besvare, og hvilke kriterier som skulle være oppfylt. Sensor E2 viste til diskusjonene som sensorgruppa hadde i forbindelse med sensureringen av besvarelsene, og hva som i fellesskap ble vektlagt av konkrete kriterier til den aktuelle oppgaven.

Ved fagprøven påpeker nemndsmedlem F4 betydningen av å bruke kriteriene inn i vurderingen og å forholde seg til dem. Andre kriterier eller momenter som lærlingen ikke er klar over skal ikke trekkes inn. Nemndsmedlem F4 viser til at alle nemndene har vært med på å diskutere vurderingskriteriene og at det derfor har blitt som det har blitt. Nemndsmedlem F1 viser til at flere vurderingskriterier burde vært mer vektlagt, noe som baserer seg på egne erfaringer i andre roller som vurderer. Spørsmålet er om det blir en del av vurderingen allikevel, og nemndsmedlem F1 sier at:

[...] jeg mener jo det at hvis det er noe som skal vektlegges så bør det stå skrevet tydelig i vurderingskriteriene. Kandidaten burde vært mer informert hvis vi skal vektlegge det mer enn vi gjør da, som jeg mener at vi skal eller burde. (F1-P15)

Nemndsmedlem F1 uttrykker at man er ikke enig om alt seg imellom i nemndene, men vedkommende har ikke problemer med å følge de kriterier som er satt opp. Nemndsmedlem F4 støtter det, og uttrykker samtidig at de kriterier som skal benyttes må være kjent for lærlingene.

Ved MECVET-testen blir bruk av kriterier preget av en mer «tvunget» form. Rater M1 uttrykker at MECVET-kriteriene kan være beskrevet noe for bredt i forhold til hva en fagperson ville tenkt på i utførelsen av sin jobb. Samtidig viser rater M3 til at MECVET-kriteriene og kriteriespesifikasjonene ikke er unormale, men de oppleves litt feil når de anvendes opp imot oppgaveteksten. Det begrunnes med at «[...] ser man isolert sett på disse kulepunktene her så har de ikke svart på det, men du sitter fortsatt igjen med en forståelse at du har faktisk klart å løse oppgaven på en relativt god måte.» (M3-P18). Det kan tolkes som at raterne vurderer på en annen måte når de ser besvarelsene i sin helhet enn når de benytter det enkelte MECVET-kriteriet og kriteriespesifikasjon. Samtidig viser rater M3 til at de første dimensjonene er viktige i forhold til selve utførelsen av faget, og de resterende dimensjonene automatisk blir en del av jobben på grunn av lover og regler.

Når det gjelder *stramhet* ved kriteriene er sensor E3 tydelig på at det ikke på forhånd må lages for faste kriterier. Det kan utelukke viktige elementer i en besvarelse man ikke kunne forutse på forhånd. Det er viktig å ivareta den reelle kunnskapen til lærlingene. Sensor E3 viser til at «[...] det nivået som kandidater har etter noen år på skole og så praktisk etterpå, så vil de vinkle det ut i fra praksis de har på det.» (E3-P12). Videre mener sensor at lærlingen må ivaretas og at kriterier ikke må være til hinder for en god vurdering. Ved åpne oppgaver gis det flere svar-muligheter, noe det må gis rom for i utforming og bruk av kriterier.

Når det gjelder fagprøven påpeker nemndsmedlem F5 at det må legges vekt på det individuelle i vurderingene, og at det er utfordrende å styre alle vurderinger innenfor noen gitte rammer. Kriteriene er med i vurderingene, men det er «[...] lærlingen som fagmann som skal vurderes og det innebærer at hele hans person må være god nok til å jobbe ute som elektriker» (F5-P15). Vedkommende sitter ikke og blar i vurderingskriteriene, og de bør ikke brukes for strengt inn i vurderingene. På spørsmål om de stemmer med virkeligheten svares det et «man synes vel det». Det kan tolkes dithen at de har noen underliggende kritiske

synspunkter på kriteriene. Ved fagprøven vil det i tillegg være en nærhet til lærlingen, noe som bringer frem spørsmålet om personlige egenskaper trekkes inn i vurderingen. Nemndsmedlem F3 uttaler at det kan skje, med blant annet lærlingenes holdninger uttrykt gjennom faglige handlinger. Det vil være utfordrende å holde seg innenfor rammen av de utviklede kriteriene. Det har paralleller til nemndsmedlem F5 sin uttalelse om å vurdere en fagmann.

Ved MECVET-testen er de stramme kriteriene allerede kommentert. Det kommer ytterligere frem ved spørsmål om det tilhørende løsningsrommet. En rater mente det kunne gi en viss grad av støtte. De resterende uttalte at løsningsrommene ga liten grad av nyttig innspill til vurderingene. Rater M4 uttaler at «Der står det mer detaljert om hvordan det skulle ha vært og da blir det nok mye mere strengere med karakterene til venstre igjen.» (M4-P18). Uttalelsen er et uttrykk for at det kan bli for snevert. Et noe annet perspektiv kom fra rater M3 som uttrykte: «Tittet litt på de løsningsrommene men jeg fikk vel egentlig aldri den, eller følte ikke jeg tok så fryktelig hensyn til løsningsrommet, fordi man sitter og vurderer dette ut i fra sin egen erfaring egentlig.» (M3-P18). Rater M3 anvender egen erfaring som et utgangspunkt for vurderingen. Det kan tolkes dithen at er stramheten for stor, kan man lett falle tilbake på sine egne erfaringer og forforståelse. Sammenhengen mellom kriteriene og det som ble beskrevet som veiledende faglige elementer i løsningsrommet var ikke tydelig.

### 5.3.3 Karakterenes betydning

Karakterenes betydning handler om forståelse og bruk. Først presenterer jeg resultater fra de utvalgte besvarelsene i hvert vurderingssystem (utvelgelse gjort rede for i kapittel 4.3.1). Ved eksamen ble det plukket ut seks besvarelser som representerte hver sin karakter, unntatt karakteren 1 (figur 33). Hensikten var å få frem forskjeller i betydningen av hver karakter. Samtidig gir det ikke mulighet for å få frem forskjeller i forståelse av samme karakter.

Besvarelse	Karakter	Bevarelse	Karakter
Nr. 1	2	Nr. 4	5
Nr. 2	6	Nr.5	2
Nr. 3	4	Nr. 6	3

Figur 33. Oversikt over karakterer fra utvalgte eksamensbesvarelser

Ved fagprøven valgte nemndene ut 11 lærlinger. Figur 34 viser nemnd 1 med besvarelse nr. 1-4, nemnd 2 besvarelser fra nr. 5-7 og nemnd 3 besvarelser fra nr. 8-11.

Besvarelse	Karakter	Bevarelse	Karakter	Bevarelse	Karakter
Nr. 1	Ikke b	Nr. 5	Bestått	Nr. 9	B meget
Nr. 2	B meget	Nr. 6	Bestått	Nr. 10	Bestått
Nr. 3	Bestått	Nr. 7	Ikke b	Nr. 11	B meget
Nr. 4	Bestått	Nr. 8	B meget		

Figur 34. Oversikt over karakterer fra utvalgte fagprøver

Det ga anledning til å få frem forståelser av hver karakter samt se på likheter og ulikheter internt på samme karakter. I nemnd 3 ble det gitt forholdsvis mange «Bestått meget godt», noe som ble poengtert at ikke var normalt.

Ved MECVET viser jeg til et samlet resultat med kompetanseprofil for alle elektroelever testet på Vg2 og alle 1. og 2. års lærlinger testet i elektrikerfaget. Dette for å gi et bilde av utviklingen over en treårsperiode. Det gis ikke noen slutt karakter i MECVET, men ratingene beregnes og sluttresultatet framkommer som en kompetanseprofil. Figur 35 viser de tre årene med elever og lærlinger.



Figur 35. Kompetanseprofiler fra elektrikerfaget på alle tre årene med elever til venstre, 1. årslærlinger i midten og 2. årslærlinger til høyre.

Det kommer tydelig fram at utviklingen av yrkeskompetansen hos elevene og lærlingene har vært lav de tre årene. Det kan fortelle oss at raterne har vært enige om den lave scoren, selv om det finnes eksempler på høyere vurderinger av enkeltelever og lærlinger.

Ved de lavere karaktererne 1 og 2 er det en kritisk grensegang mellom bestått og ikke bestått. Sensor E1 uttrykker at det mangler mye på de delene der det forventes at noe skal være besvart. Sensor E2 understøtter det og uttaler at karakteren 1 innebærer at lærlingen vil være en fare for seg selv og andre, hvor kunnskapen om el-sikkerhet ikke er tilstede. Lærlingen har åpenbart ikke evner til å til å gjøre noe særlig arbeid på elektriske anlegg. Sensor E4 uttrykker at ved karakter 1 har lærlingen noe kjennskap til forskrift eller normer, men kan ikke anvende dem. Forståelsen av karakter 1 ved eksamen kan kjennes igjen i det nemndsmedlemmene uttrykker om karakteren «ikke bestått» ved fagprøven. Ved denne karakteren uttrykker nemndsmedlemmene at mye av arbeidet som lærlingen har gjort vil være mangelfullt og ikke funksjonelt, samt ha store mangler på dokumentasjon. Et sentralt moment er el-sikkerhet rundt det tekniske anlegget og vesentlige mangler, og lite forståelse for hvor farlig strøm kan være. Rater M1 uttalte at ved laveste nivå «ikke oppfylt» vil det bety at «[...] da er det ikke nevnt eller besvarelsen er forskriftsstridig» (M1-P17).

Ved den kritiske karakteren 2 uttrykker sensor E2 at lærlingen har noe mer kunnskap og kan utføre arbeid ved enklere anlegg. Sensor E3 uttrykker følgende ved karakter 2:

En toer kjennetegner vel en som da kan gjengi deler av og bruddstykker av ting, og som for det første ikke har noen helhet i det man gjør, og for det andre da ikke kan gjengi det på en slik måte at man kan gjøre hverken hele eller deler av et arbeid, uten at noen andre står over deg og passer på deg. Her er det en løs kanon. (E3-P12)

Karakter 2 viser ifølge sensorene en delvis forståelse, men selvstendigheten mangler. Bare enklere arbeid kan gjøres og det er ingen helhet i det som gjøres. For nemndsmedlemmene vil den kritiske karakteren være på den nedre delen av karakteren «bestått». Karakteren i sin helhet ses på som en stor sekkepost. Ved den nedre delen vil det innebære at lærlingen har mangler på sitt anlegg og i dokumentasjon. Allikevel vil det kunne gi et grunnlag for å få «bestått», noe nemndsmedlem F1 uttrykker: «Men du kan gjøre en ganske dårlig fagprøve å være på grensen til stryk, og fortsatt få bestått.» (F1-P16). Rater M1 uttrykker at ved nest laveste nivå så «... har vedkommende vært innom og nevnt det.» (M1-P17). Disse uttalelsene henger sammen, og viser til begrensningene lærlingen har for å kunne gjøre en god jobb. Det er bruddstykker av en kompetanse som vises og det er ikke nok til å være en selvstendig yrkesutøver.

Ved karakterene 3 og 4 sier sensor E2 at lærlingen skal kunne utføre enkle installasjoner og være relativt selvstendig. Man kan anvende deler av kunnskapen og bruke forskrift til dels

riktig. Sensor E4 uttaler at lærlingen kjenner til forskrift og normer, og kan henviser til noe som er brukt, men alt er ikke riktig. Et sitat fra sensor E3 sier noe om både karakter 3 og 4:

En typisk 3 kandidat er jo en da som klarer å gjengi og at man klarer å unngå å gjøre de store feilene. Er ikke sikkert man klarer å lage de helt store løsningene her i verden, men man klarer ihvertfall og gjengi slik at man ikke gjør de store feilene og de elektriske anleggene ihvertfall da blir trygge. [...] Erfaringsmessig da en som oppnår en 4 på skriftlig eksamen blir stort sett en solid håndverker, og kan da såpass mye regelverk og løsninger at man da lager trygge gode og helt ok løsninger. (E3-P12)

Uttalelsen sier noe om behovet for mer selvstendighet og at mer kunnskaper trengs. Ved fagprøven uttales det: «Du kan gjøre en ålreit fagprøve og få bestått.» (F1-P16), noe som innebærer at man «[...] har en middels dokumentasjon, helt greit anlegg og grei redegjørelse.» (F1-P16). Sitatet uttrykker ikke klart hva som kreves, men det sier noe om hvor den mer generelle listen kan ligge. Dette kan ha paralleller til det som uttrykkes om karakter 3 og 4 ved eksamen. Samtidig vil karakter 4 kunne gjenspeile en god håndverker, noe som tilsvarer den øvre del av karakter «bestått». Nemndsmedlem F1 sin uttalelse om middels nivå i forhold til dokumentasjon og anlegg kan også representere en øvre del av karakteren bestått. Igjen så viser rater M1 til at på nest øverste nivå, «nesten» «[...] da er det nogenlunde godt dokumentert men med visse mangler da.» (M1-P17). Uttalelsen er mer generell, men det kan tolkes dithen at nogenlunde kan sammenlignes med middels. Samtidig uttrykker de at de savner et ekstra nivå i skalaen både ved fagprøven og MECVET-testen. Rater M4 sier: «Jeg synes det er stor forskjell på dårlig og nesten. Der synes jeg det er langt i mellom. Skulle hatt noe i mellom. Mange toere som kanskje skulle vært i midtfeltet og enere som skulle vært der.» (M4-P18). Dette har en betydning i karaktersettingen, og det kan tolkes dithen at en økende grad av nivåer kan gi mer presise vurderinger som bedre gjenspeiler kompetansnivået.

De to siste karakterene sensorene forholder seg til er 5 og 6. Lærlingen må bruke normer og forskrifter og det skal lages veldig gode løsninger som kan brukes rett til kunde. Den faglige forståelsen skal være høy. Utover det å installere utstyr og gjøre beregninger skal det også vises til fornuftige løsninger, noe lærlinger ofte gjør hvis de har vært noe ute og jobbet. Det er viktig at alle punktene er besvart, og at besvarelsen er god på hver av dem. Sensor E2 uttrykker følgende om karakteren 5 og 6:

Men en kandidat 5-6 han skal være med å kunne diskutere nye løsninger både ut i fra



et kundeperspektiv og ut fra et perspektiv som dreier seg om det samfunnsmessig og hva er vanlig, mulig og typisk, og han skal kunne se flere el.-sikkerhetsforhold og feilsøkingsforhold. (E2-P11)

Sitatet viser en stor grad av selvstendighet hvor forståelsen går langt utover kun det tekniske. Ved fagprøven skal karakteren «bestått meget godt» vise at det meste må være på plass. Løsningene skal være noe utover det ordinære og virke overbevisende. Fortsatt kan mindre deler av anlegget ikke fungere helt 100 %, og det kan være små mangler i dokumentasjonen. Men det skal ikke være noe av vesentlig betydning som ødelegger for el.-sikkerhet og for funksjonalitet av anlegg. Det skal være salgbart ovenfor kunder. De skal også kunne gi begrunnelser i forhold til forskrift og normer. Et sitat illustrerer dette:

Men er plandel super og alt ser helt supert ut nede i bua der, og de er ferdig onsdag i to tiden og alt virker og alt ser ålreit ut, og alle målinger er greit og ingen problemer med noen ting og ikke med spørsmål her oppe, da er det meget. Og er alle målinger riktig og de vet hvorfor de måler og skjønner litt bak måleresultatene, da kan de få seg en varebil på mandag morgen og kaste inn verktøyet og snu capsen bak frem, da er det meget bra. (F5-P15)

Sitatet er satt litt på spissen, men kan illustrere en helhetlig tenkning om lærlingenes besvarelser. Rater M1 sin uttalelse om høyeste nivå «oppfylt» i MECVET-skalaen, samsvarer med noe av det som ble uttalt ved fagprøven. Rater M1 uttrykker at «For at det skal være oppfylt må det være en helhet i det fra start til mål innenfor de ulike punktene». (M1-P17). Punktene i denne sammenheng kan tolkes som MECVET-kriteriene og kriteriespesifikasjonene, og disse må ses i en helhet og ha en høy score på nivået.

#### 5.3.4 Diskusjon om bruk av mål og kriterier i en helhetlig vurdering

Grunnlaget for vurdering og eventuelt karaktersetting baseres ofte på mål og kriterier. I dette avsnittet diskuterer jeg hvordan vurdererne tenker rundt helhetlig vurdering når de bruker *kompetansemål og faglige redegjørelser, vurderingskriterier og karakterer*.

Først skal vi se på hvordan vurdererne *forholder seg til* kompetansemål og MECVET-kriterier når flere av dem har trukket frem at begge disse ble sett på som vide og åpne og med et til dels for høyt nivå. Wolf (2001) sier at kommer det ikke tydelig frem hva som skal vurderes vil vurderere støtte seg på forhåndsforståelse og konsensus. Samtidig peker Dobson (2010) på betydningen av å sikre en større bevissthet om hva som ligger i mål og kriterier. En mer

rettferdig og helhetlig vurdering kan etter hans mening sikres gjennom diskusjon av mål og kriterier i fellesskap. Blant vurdererne, og spesielt ved eksamen og fagprøven, er det stor bevissthet rundt kompetansemålene, men de brukes ikke direkte i vurderingen. Samtidig peker de på at 52 kompetansemål er mye å forholde seg til, og vil være utfordrende når du skal ivareta en helhet. I tillegg var vurderer (E2) todelt i synet på kompetansemålene (noen tolket som kunnskapsmål og andre som kompetansemål), noe som kunne gjøre vurderingen enda mer utfordrende. Lahn og Nore (2015) sier at nemndene i liten grad tar utgangspunkt i kompetansemålene, men at de er med i bevisstheten under vurderingsarbeidet. Wolf (1995) mener at detaljerte beskrivelser av arbeidsdomener ikke bør gjøres. Jones (1999) hevder at det kan virke begrensende. I det lyset kan man oppfatte de vide og åpne kompetansemålene og MECVET-kriteriene som mer generelle og overordnede beskrivelser for det utøvende yrke. Ved en kompetanse-basert vurdering får man ikke målene helt tydelig beskrevet, og dermed vil ikke vurderingen kunne skje på en mekanisk måte (Wolf 1995). Derimot kan Rauner (2013) i større grad tolkes dithen at gjennom å ha et færre antall MECVET-kriterier (8 mot 52 kompetansemål) kan det skape tydeligere rammer. Utfordringen kan være at det oppleves mer som begrensende for vurdererne sett opp imot hva Jones (1999) og Wolf (1995) uttaler. Men intensjonene med MECVET-kriteriene er å fange opp de mest sentrale delene av en arbeidsprosess. Samtidig skal det legges til at Rauner (2013) mener det skal være grunnlag for tolkninger i ratingen. Det er muligheter for å kunne ha flere løsninger som kan tilpasses ulike kontekster. Selv med store forskjeller i utforming og antall kompetansemål og MECVET-kriterier, påviste raterene en sammenheng mellom MECVET-kriteriene og kompetansemålene. Vurdererne opplevde i varierende grad at det var rom for tolkninger og skjønn i vurderingssystemene og avhengig av hvordan kompetansemålene og MECVET-kriteriene rent metodisk og praktisk skulle anvendes i vurderingen. Det kan se ut til at kompetansemålene oppleves som mer åpne og fleksible og heller ikke så direkte knyttet til vurderingsarbeidet som MECVET-kriteriene.

Det er komplekse prosesser som ligger til grunn for utøvelsen av yrkene i varierende kontekster (Wolf, 1995) og grenseoppgangene på nivåene mellom elektrikerfaget og saksbehandler/tekniker/ingeniør utfordres. Studien til Tsagalidis (2008) viste at vurdererne trakk inn mer av det som ble gjort i yrket enn det som var satt opp av skolen som kriterier i vurderingene. Det kan også synliggjøre forskjeller i «nivåer». Vurderernes diskusjoner om faglige redegjørelser og grenseoppganger mellom de ulike nivåene er relevant fordi det kan ha en direkte betydning for vurderingen. Tolkningen av kompetansemål og MECVET-kriterier er

en vesentlig del i vurderingsarbeidet og blir spesielt knyttet til de skriftlige besvarelsene der innholdet i større grad diskuteres i forhold til nivåer.

Når det gjelder *vurderingskriterier* viser både Wolf (1995) og Rauner (2013) til behovet for å utvikle disse fortløpende i henhold til den konteksten vurderingene skjer i. Dobson (2010) henviser til en pendelbevegelse og jeg har valgt å se på vurderernes bruk av kriterier i de tre vurderingssystemene i forhold til et slikt bilde. Raterne i MECVET-testen kan være i den analytiske enden av pendelen med en rekke forhåndsskrevne kriterier og kriteriespesifikasjoner. Nemndsmedlemmenes bruk av kriterier ved fagprøven kan tolkes dithen at de er i den holistiske enden selv om det var en del forhåndsdefinert kriterier. Kriteriene ble sjekket ut delvis fortløpende og mest i ettertid. Hos sensorene ved eksamen er det mer en vekselvirkning mellom det analytiske og holistiske. Kriteriene og vurderingsprosessen går mer om hverandre. Jeg må legge til at distinksjonene mellom systemene ikke var helt tydelige i mitt materiale, det forekom også ulikheter internt mellom vurdererne i et vurderingssystem. Det kan knyttes til vurderernes vurderingsstrategier, noe som blir belyst i kapittel 5.4.1.

Colley & Jarvis (2007) framhever den uformelle siden av vurdering som en styrke og støtte i forhold til de formelle vurderingene. Vurderernes erfaring er en ressurs så lenge den benyttes med omhu og så lenge den kan diskuteres med andre i en sosial kontekst (Dobson, 2010). Wolf (1995) peker også på at eksperter innehar en kompetanse som kan frembringe en felles forståelse av hva faget innebærer noe vurdererne i denne studien viser. Det er vesentlig for å unngå faren med å utvikle en endeløs spiral av kriterier. Wolf (1995) presiserer at kriterier ikke kan dekke hele kompetansen, det må være rom for tolkninger. Ved MECVET-testen ble kriteriespesifikasjonene systematisk benyttet og raterene påpekte som nevnt tidligere at det kunne virke begrensende for vurderingsprosessen. Vurdererne stiller spørsmål om KOMET-modellens intensjon om å ivareta helhet i vurderingene vil slå tilbake og heller virke hemmende for helhetlige vurderinger. Det samsvarer med Wolf (2001) som påpeker at kompetansebasert vurdering ikke kan gjøres mekanisert. Det tilhørende løsningsrommet i MECVET-testen ga ifølge raterne liten støtte i vurderingsarbeidet og derfor benyttet de det i liten grad fordi de opplevde det hemmende. Samtidig påpeker vurdererne at det er en styrke at kriterier diskuteres i alle vurderingssystemene, slik Dobson (2010) og Engelsen (1998) understreker betydningen av.

Som en del av den uformelle vurderingen viser forskningen til Nore og Lahn (2015) at nemnder er en egen institusjon med egen vurderingspraksis. En mer skjønnsbasert vurdering knyttes til forventninger, kvalitet og tolkninger av det som kreves. Funnene i min studie viser at det å følge kriterier helt mekanisk ikke vil være mulig når skjønn blir en del av vurderingen. Ifølge Wolf (1995) er det både utfordrende og lite hensiktsmessig å etterstrebe en uniformert oppførsel blant vurdererne siden alle er faglige eksperter på sitt område. Analyser i HELVYRD viser at det er vanskelig for vurdererne å beskrive hvordan kriteriene anvendes. I sine studier trekker også Tsagalidis (2008) frem at vurdererne hadde kjennskap til de overordnede mål og kriterier, men at de i konkrete situasjoner hadde større oppmerksomhet på hvordan oppgavene ble løst. De studerende i hennes studie ble vurdert etter kriterier som springer ut fra yrket mer enn mål satt i skolen. Det kan trekkes linjer til hvordan vurdererne i HELVYRD bruker faglige redegjørelser knyttet til oppfatninger om hva det betyr å være en utøvende fagarbeider. Johnson (2008a) sin forskning viste at vurdererne brukte kriteriene i vurdering av portefolier og at det påvirket deres arbeid, men noen ytterligere utdyping om hvordan de ble påvirket ble ikke gjort.

Kriterier er også informasjon til lærlingene om grunnlaget for vurdering, det har derfor vært interessant å analysere arbeidet vurdererne legger i tolkning, utvikling og bruk av vurderingskriterier. Forskningen til Deichman-Sørensen et al. (2011) viser at det er en ulik grad av presisering og praksis. Resultater fra denne studien viser ulik praksis både mellom vurderingssystemene og innad i de ulike vurderingsgruppene: alt fra at det informeres om vurderingskriteriene i forkant til at lærlingene ikke blir informert om dem i det hele tatt. Det kan påvirke lærlingenes prestasjoner, samtidig som det handler om lærlingenes kjennskap til og erfaringer med de ulike prøveformene.

Flere av vurdererne presiserte at vurderingskriteriene ikke må bli for *stramme*. Det samsvarer med hva Wolf (1995) og Jones (1999) sier om kriterier som ikke må være for konkrete og konsise. Kriterier kan virke begrensende i seg selv og føre til at viktige sider av en besvarelse blir utelatt, men det kan variere fra yrke til yrke. Noe av kunnskapsinnholdet gir mindre rom for skjønn, og dermed større behov for stramhet i enkelte kriterier. I følge vurdererne i HELVYRD er det en tendens i kriterieskjemaene til både fagprøven og MECVET-testen at de blir for stramme og virker reduksjonistiske mens det i fagprøven i større grad er anledning til å gjøre en holistisk vurdering. Både Eraut (2004) og Hager og Gonczi (1994) peker på at kompetansen ikke må brytes ned i for små enheter og dermed ødelegge for helheten. Samtidig kan en større atomisering av kriteriene, som MECVET-testen kan være en representant for, gi

mulighet for å se samspillet mellom kriteriene (Hoy-Mack, 2005). På samme måte kan det oppleves som en mer utfyllende og dyptgående vurderingsprosess så lenge det gjennomføres en helhetlig vurdering til slutt. I en fagprøve mener vurdererne det er bedre anledning til en slik helhetlig vurdering av en kompetanse.

Rafoss og Witsø (2014) peker på at vurdererne har behov for større klarhet omkring kriterier og et ønske om større innflytelse på utformingen av dem. Det vil ha betydning for eierforholdet til kriteriene og hvordan de anvendes av vurdererne. Resultater fra HELVYRD viser at eierforholdene for sensorene og nemndsmedlemmene i stor grad var ivaretatt, selv om det ved fagprøven var noe mer styrt gjennom et mer statisk skjema utviklet i fellesskap. For raterene var både kriterier, kriteriespesifikasjoner og løsningsrom utviklet av andre. De opplevde det som for stramt og eierforholdet til kriteriene var i mindre grad til stede. Vurdererne stilte spørsmål om det er mulig med en helhetlig vurdering når rammene i så stor grad er styrt.

I yrker med en kompleks kompetanse skal vurdering med *karakterer* gjenspeile hva lærlingene har prestert. Vurderingssystemene i denne studien hadde tre vidt forskjellige karaktersystem. Analysen omfattet hvordan vurdererne så på bruken og betydningen av karakterer som ledd i en helhetlig vurdering. Blant annet diskuterte vurdererne om det er hensiktsmessig å ha graderte karakterer og hvor graderte de eventuelt bør være. Her kom det fram både ulemper og fordeler med færre og flere graderinger, men at det er vesentlig med tydelige grensetrekkninger i forståelsen av de ulike karakterene. Vurdererne var seg bevisst dette og diskuterte mye, slik G. S. Haaland et al. (2009) understreker betydningen av. Williams og Bateman (2003) mener at kompleksiteten i vurdering av kompetanse øker ved bruk av karakterer. Både eksamen og MECVET-testen har flere nivåer i karakterskaleringen. Karakteren som settes skal ha en verdi for lærlingen og ikke minst være valid (Johnson, 2008b). Vurdererne i studien påpeker at en for mekanisk bruk av kriterier og karakterer medføre at kompetansen som blir vurdert gjenspeiler noe annet, også i forhold til lærlingenes egen oppfatning. Williams & Bateman (2003) mener derfor det vesentlig å være fair i vurderingene og ta hensyn til kandidatene. Vurdererne kan sies å være i en maktposisjon (Hiim & Hippe, 2001), noe man må ta hensyn til i rollen som vurderer. I HELVYRD-studien kom det fram at vurdererne tok karaktersetting seriøst og diskuterte karaktergivingen grundig. Analyse av MECVET-testen og vurderernes bruk av kriterier og kriteriespesifikasjoner viste at vurdererne stilte spørsmål ved om de egentlig plasserte kandidatene litt tilfeldig på nivåene. Her viste de til det store antallet kriteriespesifikasjoner (40) det skulle vurderes etter og

hvordan det eventuelt ville svekke, forenkle og forringe kvaliteten på vurderingene. Det som fremstår som en styrke i denne studien, er at det er et to-sensorsystem i karaktersettingen i alle vurderingssystemene. Det styrker grunnlaget for helhetlige vurderinger av en kompleks yrkeskompetanse at to sensorer får diskutert besvarelsene. Dette er i tråd med hva Deichman-Sørensen et al. (2011) kom fram til i studien av prøvenemnders arbeid der de fant at det var liten grad av uoverenstemmelse om karaktersetting. I Nore og Lahns (2015) studie fant de at en ofte brukt argumentasjon ved vurdering av om en kandidat skulle bestå var om kandidaten er en som ville bli ansatt. Bruk av karakterer er sentralt i vurderinger, men må knyttes til kompetansemål og kriterier for å sikre rettferdighet i vurderingene og at karakterene kan speile rett kompetanse. I denne studien har vi sett på bruken av tre ulike vurderingsskalaer som sett i sammenheng kan være berikelse i forhold til validering av kompetanse.

### 5.3.5 Funn knyttet til helhetlig vurdering ved mål og kriterier

Vurdererne viste forståelse for og kjennskap til kompetansemålene i læreplanene, men de brukte dem ikke direkte i vurderingsprosessen. De så kompetansemålene mer som en overordnet ramme. MECVET-kriteriene som i større grad virket begrensende på vurderingsprosessen. Selv med en påvist sammenheng mellom kompetansemål (52 stk) og MECVET-kriterier (åtte stk), viser vurdererne til at mål og kriterier påvirker ulikt i en helhetlig vurdering. Det fremkom kritiske kommentarer fra vurdererne på enkelte kompetansemål og MECVET-kriterier om deres relevans.

Grenseoppgangen mellom fagarbeider og andre nivåer (tekniker/saksbehandler/ingeniør) innen fagområdet fikk betydning i vurderingene. Forståelsen av nivåene var ulik og kunne være kontekststøttet. Oftes var diskusjonen tilknyttet graden av teoretisk kunnskap lærlingen trengte, som for eksempel ved beregninger. Det har betydning for den helhetlige vurderingen. Det er et behov for en ytterligere diskusjon om grenseoppgangene.

Det var ulik utforming og anvendelse av vurderingskriterier i de tre vurderingssystemene slik jeg viser nedenfor i figur 36. Her har jeg brukt pendelmetaforen (Dobson 2010) og viser at kriteriene ved fagprøven har en mer holistisk tilnærming gjennom å være forhåndsutviklet og brukes umiddelbart, men mest i etterkant. MECVET-testen har også forhåndsutviklede kriterier, men anvendes umiddelbart og mer analytisk. Kriteriene ved eksamen kan sies å ligge midt imellom de to ytterpunktene hvor kriterieutviklingen og anvendelsen skjer fortløpende. Det er selvfølgelig nyanser her, blant annet vil det være umulig å få konkrete beskrivelser av hvilke kriterier som reelt sett anvendes ved eksamen og fagprøven. Det står i kontrast til

anvendelsen av MECVET-kriteriene og kriteriespesifikasjonene og åpner for diskusjoner om ikke kriteriene som brukes ved eksamen og fagprøven i større grad gir rom for helhetlige vurderinger.

	<b>Utforming</b>	<b>Anvendelse</b>
Eksamen	Fortløpende (dynamisk)	Mellomting (fortløpende)
Fagprøven	Forhåndsgitte (statisk)	Holistisk (i etterkant)
MECVET	Forhåndsgitte (statisk)	Analytisk (umiddelbart)

Figur 36. Oversikt på utforming og anvendelse av kriterier i de tre vurderingssystemene

Vurderernes erfaringer fra utdanning og arbeid inngikk i ulike tolkningsfellesskap med utarbeidelse av og felles forståelse av kriterier og forventninger til lærlingenes besvarelser. Her diskuterte de også grenseoppgangen mellom elektrikerens som fagarbeider og andre nivåer i yrket (tekniker/saksbehandler/ingeniør). Grenseoppgangen mellom nivåer ble også diskutert av raterne ved MECVET-testen, men de opplevde det som begrensende å anvende egne erfaringer og foreta helhetlige vurderinger på grunn av de stramme kriteriene. Vurderernes eierforhold til kriteriene opplevdes størst i de to norske systemene der de både fikk bidra i utformingen og hvordan de anvendes. Basert på mål og kriterier gir de tre vurderingssystemene varierende muligheter for å gjennomføre helhetlige vurderinger.

Måloppnåelse i form av karakterer er forskjellig i vurderingssystemene. Eksamen har en seksdelt skala (1-6), fagprøven en tredelt skala (ikke bestått, bestått, bestått mye godt) og MECVET-testen fire ulike nivåer tilknyttet 40 kriteriespesifikasjoner. Det lå en åpenhet hos vurderere for å tolke og forstå de ulike karakterene, men raterne opplevde karakterersystemet i MECVET-testen mer styrt og at det svekket helheten. Vurderernes vurderinger med karakterer var i stor grad sammenfallende, men de understreket betydningen av et to-sensorsystem med muligheter for å drøfte karaktersetning utfra en helhetlig vurdering av yrkeskompetanse. Det ble også tatt opp behov for eventuelt å ha flere eller færre nivåer i de enkelte karakterskalaene, noe som igjen vil påvirke den helhetlige vurderingen.

## 5.4 Metoder

Dette delkapitlet handler om hvordan vurdererne gjennomførte vurderingene. Det omfatter både *vurderernes strategier* for vurderingsarbeidet og hvordan de jobbet for å oppnå *samstemthet* i vurderingene.

### 5.4.1 Vurderingsstrategier

Vurderingsstrategiene knyttes til tre områder: *planlegging*, *fremgangsmåter* og *sammenligning*.

Studien viser at vurdererne som en del av *planleggingen* først og fremst setter seg inn i oppgaven. Sensor E4 sier at man ikke kan forstå oppgaven umiddelbart, og må sette seg nøye inn i den. Noen prøver å lage en løsning, mens andre setter ned punkter de mener bør være med. Sensor E4 uttrykker at man «Lager egne temaer som de må være innom for det er håpløst og ha egen fasit på hvert tema. Ingen er like!» (E4-P13). Flere leser igjennom oppgaven og en del besvarelser, og gjør seg opp en mening om hva oppgaven handler om. Sensorene møtes eller ringer hverandre for å diskutere selve oppgaven, for i størst mulig grad å skape en felles forståelse. For nemndsmedlemmene er vurderingsstrategiene preget av at fagprøven går over en uke. En del av planleggingen er knyttet til lærlingenes framleggelse av planleggingsdel. Et viktig aspekt som nemndsmedlem F2 trekker frem er det førsteinntrykket de får, og som skaper forventninger om den videre prosessen. Det sjekkes at det som skal være med er skrevet inn i planene. Planleggingsdelen til kandidaten er særs viktig, og denne skal være så godt gjennomarbeidet som mulig. Det er en forutsetning for å lykkes på fagprøven. Nemndsmedlemmene krysser av for hva som er levert av dokumentasjon, og følger deretter med på hva som er lagt til eller trukket fra underveis i uka. For nemndsmedlem F2 teller det positivt om lærlingen har gjort justeringer og endringer underveis, og at det kommer frem i lærlingens arbeid. Nemndsmedlemmene gir tilbakemelding på planleggingsdelen til lærlingene. F1 uttrykker at kommunikasjon mellom nemndsmedlemmene er en del av planlegging for oppfølging underveis. Én person greier ikke å ha like mye innsyn i alle deler av en fagprøve. Med to nemndsmedlemmer får man dekket det meste. Blant raterne sier Rater M2 at oppgaven leses godt igjennom og sier deretter hvilke kriterier som er lagt til grunn. Så går besvarelsene igjennom. Både rater M4 og M3 leser igjennom flere besvarelser som et grunnlag for å finne det riktige nivået på karakterene som skal settes senere. Begrunnelsen for det er å unngå for mye justeringer av karakterer underveis på de besvarelsene som allerede er vurdert.



Når det gjelder *fremgangsmåte* viste sensorene til at alle delene det spørres om i oppgaven skal vurderes i besvarelsene. Sensor E1 tar spørsmål for spørsmål og bruker karakterskalaen på hver av dem i den enkelte besvarelse. Til slutt blir vedkommende igjennom besvarelsen en gang til og konkluderer. Sensor E2 bruker lav-middels-høy. Først prøver vedkommende å gjøre en helhetlig vurdering, for deretter å gå inn på vurdering av hver enkelt deloppgave. Sensorene E1 og E2 kan se ut til å gå noe i motsatt vei i deres måte å gjennomføre vurderingen på. Sensor E3 leser hva lærlingen har skrevet, og er ute etter en helhet hvor sentrale faglige elementer er tatt med. Den helhetlige tenkningen kan slå ut både positivt og negativt i den endelige vurderingen. Sensorene E3 og E4 anvender tabeller i sin vurderingsprosess. Eksamensoppgaven leses igjennom først, og et excel-ark benyttes for oppgaver og deloppgaver. Sensor E4 bruker det for å holde oversikt og kjenne igjen de ulike delene. Det er bedre at lærlingene har med noe på alt, enn at noe er utelatt. Det skal komme lærlingen til gode, og svarene kan gjerne være gjemt i besvarelsen. Sensor E2 påpeker i sin vurderingsprosess hva lærlingen har lagt vekt på:

Jeg ser til og med hvilke bøker han har lest, jeg ser hvilket stoff han har brukt, jeg ser hva lærere har lagt vekt på. Så har han hatt en evne til å gjøre to ting. En å finne en løsning som i denne situasjonen kan brukes og så har han hatt en evne til å sortere bort den kunnskapen som ikke er nødvendig» (E2-P11).

Sensor E2 viser både innsikt i fagfeltet og hvilken strategi som utøves i vurderingsprosessen. Sensor E1 peker på en sentral del av vurderingsarbeidet og sier: «Problemet er jo at vi ikke ser dem, vi kjenner dem ikke og i og med at han skal sensureres ift den eksamen så må vi jo gjøre det så godt vi kan.» (E1-P10). Det skiller seg ut fra å møte lærlingen direkte.

Lik sensorene, viser også nemndsmedlemmene til betydningen av at de enkelte delene prøven består av må være med og henge sammen i en helhet (jf forskrift til opplæringsloven). Det kan ses på som en arbeidsoppgave som skal leveres til en kunde. Samtidig ser nemndsmedlemmen muligheten for at kandidater kan komme seg igjennom fagprøven med minimum på en av delene dersom det kompenseres eller veies opp av de andre delene. Men nemndsmedlem F1 påpekte viktigheten av at det må være med noe på alle delene, noe de andre nemndene støttet. Nemndsmedlem F1 uttrykker at det ikke er noen offisiell vektlegging av de fire delene i fagprøven. Men for vedkommende er det planlegging, gjennomføring og dokumentasjon som er likt vektet. De ses på som naturlig å ha med i forhold til en arbeidssituasjon.

Egenvurderingen oppleves som mer utfordrende å vekte inn i vurderingen i sammen med de

andre. Strategien og metoden for gjennomføring av vurderingen underveis er stort sett lik blant nemndene. Men det er nyanser knyttet til hvor ofte de er innom lærlingene og hvordan de forholder seg til lærlingene. Det å snakke med lærlingene og samtidig observere deres utførelse var vanlig i alle nemndene. Nemndsmedlem F2 uttrykker at det kan stilles direkte spørsmål til en lærling underveis, noe som kan overraske vedkommende, siden selve eksamineringen er flere dager senere. Men lærlingen er til gjenstand for vurdering under hele fagprøven. Det kan gi nemndsmedlem og lærling en pekepinn om det arbeidet som pågår og skape forutsigbarhet for lærlingen frem til den dagen de skal presentere og forsvare arbeidet sitt. Nemndsmedlem F3 trekker frem viktigheten av å stille spørsmål underveis for å få lærlingene til å reflektere og lære underveis. I tillegg til å stille spørsmål uttrykte nemndsmedlem F5 viktigheten av å følge med på lærlingene hele veien, for eventuelt å fange opp ting som det kan veiledes på. Et eksempel er nemndsmedlem F2 som viser til veiledning av lærlinger på bruk og tolkning av oppslagsverk, og gjennomføring av en sluttkontroll. Enkelte lærlinger kunne lære det de ble veiledet på underveis og anvende det i fagprøven, noe som gledet nemndsmedlemmet. Et annet eksempel kommer fra nemndsmedlem F4. Lærlingene får gjøre feilsøking i anleggene når de ikke fungerer, hvor de virkelig får vist sin kompetanse. Synet på lærlingene og den situasjonen de er i framkommer gjennom nemndsmedlem F1 sin uttalelse «Jeg føler ihvertfall det at vi må ta litt høyde for at de kan mer enn det de klarer å få frem og pga de situasjonene de er i. Det mener ihvertfall jeg da.» (F1-P16). Dette er et vesentlig aspekt som ivaretar lærlingene i den konteksten de er i og som tar hensyn til deres forutsetninger.

Raterne uttrykker at en helhetlig vurdering innebærer å legge vekt på hele arbeidsprosessen fra planlegging til ferdig produkt. Rater M2 sier man må se på hele arbeidsoppgaven og ikke bare en planleggingsdel eller gjennomføringsdel. Det må ses på som et produkt du skal levere fra deg. Rater M1 viser det samme ved å se etter faglige begrunnelser, og om forslagene er egnet til forutsatt bruk og besvarer kundens behov. Rater M3 viser til sin prosess rundt vurdering av MECVET oppgavene, og sammenligner med vurdering i andre prøveformer. Det beskrives som en normal ting å prøve å skaffe seg «[...] den helhetlige følelsen hvor du på en måte leser igjennom oppgaven og leser igjennom svaret, så på en måte danner du deg et bilde på om du har svart oppgaven eller ikke.» (M3-P18). En mer overordnet vurdering tas etter å ha lest igjennom hele besvarelsen. Ved MECVET opplever Rater M3 å være mer styrt av kriteriene i vurderingsskjemaet og uttrykker dette slik: «Så leser jeg igjennom og så egentlig så svarer jeg på vegne av han disse spørsmålene så det er jo sånn jeg har vurdert dette her.»

(M3-P18). En overordnet karakter blir det ikke tatt hensyn til, men man svarer på de ulike kriteriespesifikasjonene og setter et nivå. Samtidig prøver Rater M3 å finne ut om den vil fungere i virkeligheten gjennom å visualisere lærlingens besvarelse, og vurderes den opp imot hva kunden har bestilt. Rater M3 uttrykker at « [...] oppgavene er på en måte en situasjonsbeskrivelse som det er gjort her og det finnes ikke noe fasit og det kan gjøres på mange forskjellige måter.» (M3-P18). Rater ser oppgaven som åpen, og selv med mer styrt gjennomføring, uttrykkes et behov for å se det mer helhetlig. Følgende sitat underbygger det: «For det at du sitter der med en forståelse i et perspektiv om at dette har han virkelig peiling på. Det handler om egentlig hvilke følelser du får da.» (M3-P18). Dette viser at det er et rom for å trekke inn mer helhetlig tenkning ved rating også. Rater M4 viser til noe av den samme strategien. En del av gjennomføringen er å lese igjennom hele besvarelsen for å få et inntrykk av om denne lærlingen har skjønnet oppgaven. I tillegg kommer det frem at en del av strategien er å sammenligne opp imot seg selv og kollegaer rundt på egen arbeidsplass, både blant de som fungerer godt og de som ikke fungerer så godt. Det utgjør en referanseramme for rater M4 med tanke på å se helheten i besvarelsen.

Når det gjelder *sammenligninger* av besvarelser forekom dette i alle vurderingssystemene. Sensor E4 uttrykte at det hender de gjør sammenligninger og spesielt ved karakteren 1 og 2, som er et kritisk punkt med hensyn til å gjøre de riktige vurderingene. Det innebærer blant annet at man ser besvarelser opp imot hverandre. Det samme gjorde nemndene ved fagprøven der de sammenlignet lærlingers arbeid, men stort sett var dette innenfor den samme karakteren og spesielt karakteren «bestått». Samtidig ble det presisert fra en nemnd at det ikke er anledning til å gjøre så mye sammenligninger, da det skal settes en individuell karakter der og da. Ved MECVET forekom det også sammenligninger, noe raterne begrunnet med de åpne oppgavene der det ikke er noe fasitsvar og dermed mente de sammenligninger var nødvendig innimellom.

#### 5.4.2 Vurderernes samstemthet

Samstemthet handler om vurderernes *enighet* og *uenighet* i vurderinger og karaktersetting.

Flere av sensorene viste til en første sortering av besvarelsene i forhold til lav, middels og høy karakter. I denne fasen ble det oppfattet som stor grad av *enighet*. Sensor E2 uttrykker at man blir enig med medsensor om hvor man legger tyngdepunktet på vurderingene. Sensorene har først en generell diskusjon om faglige deler de skal fokusere på. For å komme til enighet uttrykte sensor E1 at det kunne være lurt å legge seg litt taktisk på karakterskalaen for lettere å

kunne justere opp eller ned. Dette var i situasjoner hvor sensor selv kunne være i tvil før samsensurering. Samtidig forklarer sensor E1 at det er stor grad av enighet med medsensor ettersom de har samarbeidet i mange år. Blant nemndsmedlemmene var det også stor grad av enighet om karakter, men det fremkom også uenigheter på enkelte faglige områder. Ved tvilstilfeller ble vurderingen nøyere gjennomgått. Et tilfelle viste tvil rundt karakteren «bestått» hvor nemndsmedlem F1 uttrykker at «Men jeg tror fortsatt det var riktig å la han få «bestått» og det virket som han ikke hadde gjort noen sikkerhetsmessige feil, og det virket som han hadde en forståelse for det han drev med og.» (F1-P16). Usikkerheten hadde vært større om lærlingen hadde fått stryk, selv om denne lærlingen hadde store utfordringer med gjennomføringsdel og produktet som skulle leveres. Lærlingen ga selv uttrykk for at stryk kunne forekomme, men allikevel ga nemnda karakteren «bestått» basert på opplevelsen av lærlingens innsikt ved egenvurderingen. Flere faktorer spilte inn og utfordret nemndsmedlemmene ved den endelige karaktersettingen, men en helhetlig tenkning ble benyttet i dette tvilstilfellet. Et annet eksempel var en diskusjon om «bestått» eller «bestått meget godt», hvor nemndsmedlem F5 uttaler at det ble karakteren «bestått meget godt», selv om selve gjennomføringsdelen ikke var helt topp. For å vippe ned, måtte lærlingen ha hatt problemer i den muntlige delen. Den helhetlige vurderingen ga lærlingen Meget «bestått», noe som kan vise at det er et spenn i Meget-karakteren også. Nemndsmedlem F1 mener at er man i tvil så skal det komme lærlingen til gode fordi man som nemndsmedlem kan ta feil. Raterene ved MECVET uttrykte i mindre grad tvil om vurderingene, men man kunne ha noe ulik oppfatning om tolkning av enkelte kriteriespesifikasjoner. Gjennom diskusjoner og omforente vurderinger kom raterne fort fram til felles forståelse og enighet, ingen ratere uttrykte noen direkte tvil på hvilket nivå de skulle plassere de ulike kandidatene. Think-aloud-lydopptakene bekreftet at det i ratingprosessen var stor grad av enighet blant raterne. Svarene kom raskt på valg av karakter, men begrunnelsene var i mindre grad utfyllende.

*Uenighet* opplevdes i liten grad, men det var noen eksempler som ved en eksamensbesvarelse der det var sprik på to karakterer mellom sensor E1 og E3. I dette tilfelle var det kritisk fordi det var snakk om stryk eller ikke stryk. Sensor E1 hadde karakteren en og den andre hadde karakteren tre. Karakteren etter samsensurering endte på to. Sensor E1 uttalte at ved fare for stryk, så vurderes det alltid nøye. Stryk var sannsynlig og sensor E1 sa at «Hvis han ikke hadde levert en tegning så hadde han «ikke bestått», selvom tegningen var dårlig ...» (E1-P10). Samtidig uttaler sensor E1 senere i intervjurunde to at hadde vedkommende vurdert alene, så ville lærlingen fått karakteren to. Sensor E1 viste til nytten å ha en medsensor å

diskutere med og uttalte at «[...] kanskje jeg var litt streng.» (E1-P10). Det presiseres tydelig viktigheten av å vurdere ekstra nøye når det er snakk om stryk, og også fordelene med å være to stykker. Sensoren E3 begrunnet sin karakter med at den inneholdt en del basis-ting med forklaringer og

[...] som har vært savnet ved tidligere års eksamener da hvor de ikke gjør noen spesielle forklaringer i det hele tatt, men bare står et regnestykke. Så det er det som jeg synes trekker den opp da, men jeg ga meg i diskusjonene og man møtes i midten» (E3-P12).

I utgangspunktet fokuserte sensorene på litt forskjellig innhold. Sensor E3 viste til diskusjoner i sensorgruppa om viktigheten av lærlingenes begrunnelser for valg og sensor E1 hadde fokus på det med tegninger. Allikevel møttes de på midten og ble omforent. Sensor E1 viser til at usikkerhet og forskjeller i vurderingene kan knyttes til uklarheter i oppgavetekstene. Dette kunne medføre strengere sensur og større forventninger til lærlingene i deres besvarelser. Et annet eksempel er to sensorer med henholdsvis sterk tre og svak fire. Dette er ingen stor forskjell, men sluttresultatet endte opp med karakteren tre, noe som i realiteten betyr at begge gikk ned. I fellesskap så sensorene på hva som var gjort bra og det som ble sett på som store hull og mangler i besvarelsen. Sensor E4 viste til at runden med medsensor ga et tydeligere bilde av hvilken karakter som var riktig, og at karakteren tre ble den naturlige å velge. Sensor følte ikke dette gikk på bekostning av egen oppfatning, han sa det var greit og uttrykte at «Jada, jeg tok ikke tak i spiserøret på han, hehe.» (E4-P13).

Ved vurdering av fagprøven kom det ikke til uttrykk noen uenighet om karakter. Det ble diskutert tvilstilfeller, men da hadde som oftest begge nemndsmedlemmene tvilt. Det kan også være variasjoner innenfor samme karakter, spesielt er det karakteren «bestått» som gir rom for flere synspunkter. Et nemndsmedlem viste til at i vedkommendes tid som vurderer av fagprøver har det aldri oppstått uenigheter. Blant ratere var det også i mindre grad uenigheter, men på grunn av skaleringen og skjemaet som brukes i vurderingsprosessen var det forskjeller, som oftest bare ett nivå. Var forskjellene større, havnet vurderingen som oftest på karakteren i midten. Ved en uenighet handler mye om å bevisstgjøre hverandre, siden oppfatningene kan være ulike omkring de samme besvarelsene eller kriteriene. Jeg gjengir et sitat som kan vise et eksempel.

- Langsiktig bruk utvidelse og fornyelse ... 0.M3

- Oi! Der har jeg en toer. Jeg har nok sett i forhold til utvidelser han skriver noe om, samt skjult anlegg i en vegg her etter at panel er borte.M4
- Jeg har nok sett den isolert i forhold til om han har sagt noe til å ha muligheter for å kunne å utvide anlegget med flere vern i sikringsskapet og sånne ting kanskje.M3
- Det har han jo ikke gjort da.M4
- Og det spørs jo bare hvordan du legger listen hen.M3
- Vi går ned til en ener da. Du opp.M4

(M3, M4-P18)

I utgangspunktet var forskjellen på to nivåer, men gjennom en kort diskusjon og bevisstgjøringsprosess kom raterene til enighet. Dette var typisk for situasjonene der raterne i utgangspunktet var uenige.

#### 5.4.3 Diskusjon av vurderernes metodebruk i helhetlige vurderinger

Den påfølgende diskusjonen vil være vurderernes metodiske anvendelser i vurderingsprosessen med *vurderingsstrategier* og *vurderernes samstemthet*.

*Planlegging* er første ledd i vurderernes vurderingsstrategier og her viste de til at de starter med å sikre seg en god forståelse av oppgaven og hva som skal vurderes. Det igjen gir grunnlag for å planlegge hva og hvordan vurderingene skal gjennomføres. Kanskje vil det være aktuelt med flere vurderingsmetoder slik O'Grady og Alwis (2012) anbefaler når prosessene rundt vurdering er komplekse og det trengs ulik informasjon som kan bidra til en god vurdering. Kompleksiteten krever også at vurdererne er bevisste på at noen vurderingsformer kan være for snevre til å avdekke den reelle kompetansen. Eisner (1991) viser til behovet for å forholde seg til flere kilder og ulike sanser, blant annet kan hans begrep «connoisseururship» være en del av planleggingen spesielt for nemndsmedlemmene ved fagprøven. De kan se sammenhengen med både prosess og produkt i vurderingen, hvor både formativ og summativ vurdering blir trukket inn. Eisner viser til at hva vi bedømmer og trekker inn i vurderingene er en vesentlig del av planleggingen. Dette kan sees i sammenheng med hva Dobson (2010) skriver om helhetsvurdering som kan åpne utover og kriterier som kan snevre innover. Crisp (2012) viser til viktigheten av å sette seg inn i studentenes arbeid og planlegge vurderingsarbeidet for å få et innblikk i hvordan studenten har realisert oppgaven både i forhold til lengde og fokus, noe vurdererne i denne studien også var opptatt av.

Helhetlig vurdering i yrkesfag får konsekvenser for *fremgangsmåter* i vurderingsarbeidet (Dobson, 2010). Det har paralleller til Wolfs (1995) uttalelse om å være «vocational» i vurderingen av en profesjonell og arbeidsbasert kompetanse. Vurderernes egen faglige standard vil være utgangspunkt for vurdererne og ikke en mekanisk benyttelse av et sett med forhåndsdefinerte kriterier (Wolf, 1995). Deichman-Sørensen et al. (2011) viser til at det er den faglige forståelsen og helhetlige kompetansen som skal legges til grunn i vurderingsarbeidet, inkludert det skjønnsbaserte. Selv med de ulike fremgangsmåtene vurdererne i HELVYRD viste til, kan det sies at både ved eksamen og fagprøven signaliserte vurdererne behov for å se lærlingenes besvarelse i en helhet. Det gjaldt også skriftlig eksamens med deloppgave-vurdering. Hoy-Mack (2005) nevner samspillet mellom det analytiske og det helhetlige, noe Dobson (2010) også påpeker. I fagprøven er det mulig å være mer «vocational» i vurderingen. Her er muligheten for kommunikasjon og veiledning med lærlingene underveis og det er en praktisk utførelse som gir viktige innspill til helhetlige vurderinger slik Dobson (2010) og Wolf (1995) snakker om.

Vurderingsstrategien ved MECVET ser ut til å være mer styrt, selv med mulighet for enkeltoppgave-lesning og fleroppgave lesning. Vurderingsprosessen ble av vurdererne opplevd mer mekanisk når de fulgte kriteriespesifikasjonene, noe som også påvirket deres muligheter for helhetlige tenkning og vurdering. Sadler (2009) viser til at summen av kriterier og et helhetlig perspektiv kan ende i to vidt forskjellige resultater, noe vi så hos raterene i MECVET-testen. Deres helhetlige opplevelse av besvarelsene ga ulikt resultat med og uten bruk av kriteriespesifikasjonene. Og peker på hvilken betydning fremgangsmåten i vurderingsarbeidet har. Den styrte strategien kan oppleves å stå i kontrast til at MECVET har åpne oppgaver med mulighet for flere løsninger. Det finnes tolkningsmuligheter innenfor de enkelte kriteriespesifikasjonene som kan underbygge et helhetlig syn, men det må rettes et kritisk blikk på mekanisk summering av alle kriteriene og om vurderingene blir gjort på en «vocational» måte. Sadler (2009) ser behov for en mer åpen vei for å kunne gjøre helhetlige vurderinger der kriterier ikke må være en ferdig liste. «Connoisseurship» verdsetter (appreciative) det man ser og bruker det på en «*fine-grained*» måte i vurderingen (Eisner, 1991). Analyser av data fra vurdering ved eksamen og fagprøver viser at fremgangsmetodene i vurderingsprosessen er mer åpne. Kommunikasjonen med lærlingene og praksisgjennomføring ved fagprøven ga viktige bidrag til «*fine-grained*» vurdering. Crisp (2012) forskning viste til vurderernes emosjonelle og sosiale reaksjoner ved vurderingen, alt fra glede til det å mislike det de leste. Også i HELVYRD-studien viste vurdererne

følelsesmessig engasjement og innlevelse. Crisp (2012) sier også at vurdererne skal kunne «prate» med studenten slik at vurdering foregår både underveis eller samtidig som besvarelser gjennomgås før det til slutt gjøres en overordnet og endelig vurdering.

I studien har vi også sett at det i alle vurderingssystemene vil være naturlig med *sammenligninger* mellom oppgaver og besvarelser, slik flere forskere omtaler (Vaughen, 1991; Olsen, 2009; Crisp, 2012). Vurdererne påpeker at mange besvarelser kan være likeverdige på vurderingsnivå og ha samme karakter, men fortsatt være forskjellige. Det viser kompleksiteten i å vurdere yrkeskompetanse, og utfordringer i det å gjøre sammenligninger. Vi har også sett at åpenheten i oppgavene med mulighet for flere løsninger gjør at flere vurderere foretar sammenligninger og ser det som viktige innspill i den helhetlige vurderingen og at det kan være med på å sikre en større rettferdighet i vurderingene. Grimen og Molander (2008) viser til at det også kan være problematisk å se på tidligere vurderinger av oppgaver og besvarelser fordi de ikke er helt deskriptive (de er også normative og verdibaserte). Selv om sammenligninger var fremtredende i vurderingene i mitt materiale, var det ingen tvil om at det var den kriterie-baserte vurderingen som lå til grunn. Et eksempel på dette var antallet meget bestått blant lærlingene på fagprøven, noe som ble sett på som en «unormal» fordeling av karakterer.

Materialet viste en stor grad av *enighet* mellom vurdererne når de ga den endelige vurderingen. Vurderernes *samstemthet* er interessant sett i lys av et to-sensorsystem og muligheter for å sammenligne resultatene mellom de ulike vurderingssystemene. Tidligere forskning på prøvenemnders vurderinger viser også at det i liten grad av uoverensstemmelser (Deichman-Sørensen et al. ,2011). Vaughen (1991) uttrykker at graden av pålitelighet ved helhetlig vurdering er høy, selv etter en stund, noe som kan være med på å bygge opp under tendensen mot enighet. Samtidig viser Grimen og Molander (2008) at i forbindelse med besvarelser uten fasit svar, som alle disse vurderingssystemene representerer, er enigheten som oftest stor, men skjønnnet kan true den likebehandlingen man etterstreber. Av natur kan skjønn som vurderingsform gi utfordringer hvor konsistenskrav kan være vanskelig å tilfredsstillere. Samtidig vil den enkeltes utøvende skjønn kunne fanges opp og bli utfordret i diskusjoner med medsensor og være en påvirkende faktor i det å være samstemt. I HELVYRD har vi også sett diskusjoner som inkluderte flere vurderere enn de to sensorene, for eksempel ved eksamen. Strategiene for å sikre enighet ser ut til å ligge i lang erfaring som vurderer, felles diskusjoner og forståelse av oppgaver og kriterier og karakterskaleringer. Materialet viser at vurderernes diskusjoner rundt ulike besvarelser er en sentral del av



helhetlige vurderinger. I komplekse vurderingsprosesser vil bevisstgjøring gjennom det å ha flere øyne på de samme besvarelsene være sentralt. Enkelte av resultatene hadde tilsynelatende vært annerledes om vurdererne hadde vurdert alene.

Analysene viste at det også var *uenighet mellom vurdererne*, spesielt når det gjaldt deler av innholdet. Grimen og Molander (2008) uttrykker at ved individualiseringsprinsippet må man regne med en fornuftig uenighet, og at konklusjonene kan bli noe ulike, da det er flere kilder vurdererne støtter seg til. Samtidig kan kildene i form av uklare oppgaver og kriterier være en utfordring som balanseres gjennom et to-sensorsystem. I den sammenheng kan man stille spørsmål ved hvor dedikert den enkelte vurderer er til å få gjennomslag for sine meninger når konsensus er målet. I HELVYRD-studien var det ingen tvil om at vurdererne gjorde grundige begrunnelser for sine valg ved tvilstilfeller med fare for stryk. Grimen og Molander (2008) viser til at forskjeller også kan oppstå ved førstepersonserfaringer. Ingen vurderere kan erstatte hverandre uten å være klar over at det kan bli forskjeller i resultatene. Dette viser igjen betydningen av å ha et to-sensorsystem som kan fange opp og hindre uheldige utslag av individuelle forskjeller.

#### 5.4.4 Funn knyttet til helhetlig vurdering ved metoder

Planlegging gjorde vurdererne ved å sette seg godt inn i oppgaver og besvarelser. Spesielt gjaldt det bevissthet om oppgavens muligheter og begrensninger ved eksamen og MECVET-testen, noe som ifølge Eisner (1991), O`Grady og Alwis (2012) og Dobson (2010) er vesentlig for helhetlig tenkning. Raternes planlegging ble påvirket av MECVET-testens stramme rammer med ferdigutviklede skjemaer. Fagprøven skilte seg ut med at førsteinntrykket av lærlingenes arbeid med planleggingsdelen var utgangspunkt for nemndsmedlemmenes videre planlegging av vurderingsarbeidet. Eksamen hadde et større handlingsrom i planleggingen, men sentralt var sensorenes samarbeid i vurderingsprosessen. Ulike strategier ble brukt av vurdererne for å se helhetlig på besvarelsene. Sensorene ved eksamen gjennomførte både deloppgave-vurdering og vurdering med fordypning, men helheten ble alltid vurdert til slutt. Bruk av tabeller og skjemaer var en del av strategien for å systematisere vurderingen og for å sikre at alle delene var med i besvarelsen. MECVET-testene styrte i større grad vurderernes fremgangsmåter samtidig som raterne etterstrebet å se en helhet gjennom å visualisere om løsningen fungerte. De påpekte at vurderingene ville bli forskjellige avhengig om man tenkte helhetlig uten å forholde seg til kriteriene kontra en systematisk vurdering basert på kriteriespesifikasjonene. Nemndsmedlemmene hadde et større grunnlag for å tenke helhetlig.

De kunne være mer «vocational» både ved å kommunisere og observere lærlingen gjennom den praktiske delen. Som en del av vurderingene viste vurdererne engasjement og innlevelse og tok lærlingenes besvarelser på alvor, noe som samsvarer med forskningen til Crisp (2012).

I alle vurderingssystemene ble besvarelsene sammenlignet som ledd i å sikre en rettferdig og riktig vurdering, spesielt med åpne oppgaver og flere løsningsalternativer. Det handlet om å skaffe seg forståelse og innsikt fra ulike vinklinger for å øke bevisstheten rundt helhetlige vurderinger av enkeltbesvarelser.

Vurdererne var i stor grad samstemte i vurderingene samtidig som det kunne være ulikt syn på faglig innhold. Lang erfaring, tett samarbeid med andre vurderere over mange år og samsensurering i to-sensorsystem bidro til konsensus basert på helhetlige vurderinger og skjønnsutøvelse. Både individualiseringsprinsippet og førstepersonserfaringer (Grimen og Molander 2008) viser betydningen av diskusjoner og styrken ved to-sensorsystemet i helhetlig vurdering av kompetanse, noe vi så i alle de tre vurderingssystemene.

## 5.5 Forutsetninger

Vurdering av yrkeskompetanse preges også av vurderernes bakgrunn og forutsetninger for vurdering. Dett delkapitlet handler derfor om *fagpersonen som vurderer*.

### 5.5.1 Fagperson som vurderer

Vurdererne som fagpersoner knyttes til tre områder: *faglig bakgrunn, skole og arbeidsliv og yrkesliv*.

Om *faglig bakgrunn* uttrykker sensor E4, som er lærer, at det ikke er tvil om hva som ligger til grunn for vurderingene: «Det er jo ut fra mine erfaringer, planlegge hva jeg mener må være med da. Det er min faglige bakgrunn i bånd da. Helt klart!» (E4-P13). Uavhengig av om den som vurderer er en representant fra skole eller arbeidsliv, uttrykker vurdererne i HELVYRD at deres egen faglige bakgrunn er vesentlig i vurderingene. Nemndsmedlem F3 viser til en opplevelse hvor vedkommende fikk inntrykk av at lærlingen kunne mye mer utfra handlinger under fagprøven, men opplevde det som en stor utfordring å få frem noe muntlig fra lærlingen som kunne underbygge antakelsene. Nemndsmedlemmet forstår intuitivt faglige situasjoner, men det er ikke så lett å uttrykke det eksplisitt. F3 uttaler:

Er litt sånn som når du går og gjør en jobb som fagarbeider, så går du og kaster over et blikk og du ser mye med en gang og du har en kompetanse på det, og det tenker jeg litt

her og. Du ser hvordan de jobber og du gjør deg noen tanker så kaster du et blikk på det, men må jo gå inn i det for at det ikke skal bli overfladisk. Men du får en sånn det her går bra. Er ikke ofte du tenker at det har gått dårlig altså. Det stemmer som regel det, men jeg tenker at det handler om at vi har den fagkompetansen selv. (F3-P14)

Sensor E3 som også er nemndsmedlem og vurderer fagprøver viser til de ulike rollene ved eksamen og fagprøven. Ved skriftlig eksamen er lærlingene anonyme, mens det ved fagprøven blir en relasjon til lærlingene og det dannes et førsteinntrykk. Gjør lærlingen et godt førsteinntrykk, eller - motsatt tilfelle - dårlig, tas det med i vurderingen enten du vil eller ikke. Samtidig uttrykker E3 at det er lettere å forholde seg til kun kompetansemål ved skriftlig eksamen, uten kjennskap til lærlingen. Den nærheten nemndsmedlemmene har til lærlingene viser at de i sin rolle som vurderere bruker sin faglige bakgrunn for å etterstrebe å få frem de positive sidene i lærlingenes arbeid (spesielt ved faren for stryk). Samtidig var de tydelige på å presisere hvilke sider som var mindre bra. Nemndsmedlem F4 opplever det som viktig å stille lærlingene gode spørsmål. Vedkommende vil ikke opptre som en person som er ute etter å «ta» lærlingene, men heller gi dem inntrykk av at de ønsker å få dem igjennom fagprøven. Nemndsmedlem F2 uttrykker at ikke alle lærlingene skal gå inn i en A4-form, fordi mennesker er forskjellige. Nemndsmedlem F5 er også opptatt av å gi lærlingene en positiv opplevelse og ønsker ikke at de skal føle seg krenket. Følgende uttrykkes:

Så er det jo noen som er full av nerver, og jeg gikk under navnet strykejernet i mange år. Var vel noen som mente jeg strøk mange, men alltid fått beskjed om at jeg er rettferdig da og skryt av det. Det er for å danne meg et bilde av dem og er det noen som er fryktelig nervøse så setter jeg med bordet og prater litt med dem, og forteller litt hva vi forventer og hvor vi synes de er, og hva de har igjen og må legge vekt på for å komme videre. Prøver hele tiden og roe dem. (F5-P15)

Nemndsmedlem F5 viser her forståelse av egen rolle som fagperson og vurderer for å ivareta lærlingen i vurderingsprosessen gjennom fagprøven. Rater M1 sier helt konkret at det er som fagperson det reageres på en del ulike tekniske løsninger som lærlingene kommer frem til i sine besvarelser. Rater M2 viser til betydningen av å kunne sitt fag når spørsmålet om relevans til de åtte MECVET-kriteriene skal vurderes. «Er du dreven på fag og vurderer så tenker jeg at du kan faget og det er kanskje det som er litt unikt. Kan du faget godt så vet du hvilket nivå du skal legge det på.» (M2-P17). THer er det tydelig at det å inneha en fagkompetanse har betydning for vurderingene.

Ofte stilles det spørsmål om det betyr noe om vurderere kommer fra *skole eller arbeidsliv*, om det har noen innflytelse på vurderingene. Sensor E3, som representerer en bedrift, synes det er interessant med en sensorgruppe som kan lære av hverandre. Vedkommende lærer veldig mye om skole og hvordan lærere jobber og tenker, og håper samtidig at lærere og skole kan lære mye av hvordan en bedrift jobber og tenker. Som fagperson og vurderer med bakgrunn i en bedrift opplever vedkommende å være mer opptatt av helhet i besvarelsene og ikke så mye detaljer. Sensor E3 får inntrykk av at enkelte lærere er mer opptatt av detaljene og at disse må være på plass for å få en god vurdering, men presiserer at lærerene kan være uenig i denne uttalelsen. Ved fagprøven viser nemndsmedlem F5 til sin lange erfaring i yrket og uttrykker at det er dette vedkommende kan. I forlengelsen stilles spørsmålet om det er vedkommende som fagmann som gjør disse vurderingene. Svaret er: «Ja, det er helt klart for jeg lurer på, kanskje galt av meg, hvis skoleverket skal gjøre dette her som har en annen vurdering av det utelivet som skjer ute på byen hva vi får ut da.» (F5-P15)

I denne delen av HELVYRD har jeg sett på vurdererens faglige bakgrunn og hvilken betydning det har for vurderingene. Hva betyr det at representanter fra arbeidslivet vurderer en eksamen eller en lærer vurderer fagprøven? Vurdererne diskuterer ofte forholdet mellom skole og arbeidsliv og hvordan samarbeid mellom lærere og fagpersoner kan være fruktbart. Rater M2 trekker frem sin rolle som yrkesfaglærer, og vedkommende mente dette fikk betydning i vurderingsprosessen i form av å se mer positivt på en besvarelse, noe som kan medføre bedre karakter. Samtidig viser rater M2 til betydningen av å kjenne til den eksisterende teknologien som brukes ute på anlegg, og at det finnes lærere i skolen i dag som ikke har vært ute i yrket og skjønner dette. I vurdering vil det være en utfordring å benytte lærere som ikke har oppdatert og relevant kompetanse. Rater M4 er en utøvende elektriker og opplever de andre raterne som mer erfarne og skolerte i det å jobbe med vurdering. Vedkommende synes det er nyttig og spennende å høre hvordan de andre tenker om de lærlingene vedkommende til daglig jobber med i sitt virke. Rater M4 peker på hvordan krav til vurderingskompetanse kan ivaretas i rollen som vurderer.

Det siste området handler om hvordan vurderernes erfaringer med *yrkeslivet* kommer til uttrykk i vurderingene. Sensor E1 viser til en svært god besvarelse og uttrykker det slik «[...] du får en sånn som er så gjennomarbeidet som en slik besvarelse er så må han ha praktisk erfaring. Ellers så hadde han ikke klart å få til denne besvarelsen som han har.» (E1-P10). På spørsmål om hvordan man kunne lese det ut av en skriftlig besvarelse uttrykker sensor E1 at det er vanskelig, men man kan greie å forstå det med denne type besvarelser. Sensor E1 sier

det finnes eksempler på lærlinger som er ferdig utdannet elektrikere før de har tatt sin egen fagprøve. Samme sensor trekker frem eksempel fra yrkeslivet på at det nesten ikke er noen elektrikere som tegner eller beregner i sitt daglige virke, og dermed vil de lærlingene lide under vurderingene av disse eksamensoppgavene. Sensor E2 viser til et annet eksempel om hvordan en lærling kunne fungert i elektrikerfaget:

Jeg er ikke tvil om at han kunne vært med en elektriker og gjort den installasjonen og forstått rådene og instruksjonene han hadde fått, men han ville ikke kunne gjort dette selvstendig og det er det ingenting i papirene som viser ihvertfall.» (E2-P11).

Sensor viser tydelig den aktive bruken av yrkeslivet for å analysere lærlingens besvarelse, hvor det til slutt må tas stilling til følgende: «[...] kan denne kandidaten på en sikker måte være med på å jobbe på et elektrisk anlegg og hvis svaret på det er ja så er det 2. Er svaret nei så er det stryk altså.» (E2-P11). Det kan tolkes dithen at sensor viser til både selvstendighet i utførelsen av yrket og om lærlingen vil passe til yrket. Det skal gjennomføres en helhetlig jobb med et oppdrag og en levert dokumentasjon til slutt, noe som utgjør grunnlaget for den helhetlige vurderingen. Sensor E3 uttaler at «Jeg ønsker hele tiden å ha i bakhodet sluttkompetanse som er fagbrev, og det her skal føre frem til fagbrevet og den helheten jeg ønsker å se, og da vil blant annet praktiske løsninger være positivt.» (E3-P12). Spørsmålet stilles om lærlingen har det som skal til for å bli en elektriker, så helhetlig vurdering blir et moment i vurderingen.

Nemndsmedlem F2 viste til sin erfaring som elektriker og hvordan vurdering av kollegaer var en viktig del av arbeidshverdagen for å få gjennomført ulike arbeidsoppdrag. Den erfaringen trekker vedkommende med seg inn i vurderingsarbeidet ved fagprøven.

Sånn opp igjen et langt liv i elektrikerfaget kan jeg godt vurdere kollegaer til om de kan bli satt til det dem skal gjøre når du har ansvar for større bygg. Den tankegangen har jeg dratt med meg inn i prøvenemnda og arbeid, og det er ikke gitt at en elektrikerkollega kan ta ansvar for en del av bygget. Det er en helhetlig vurdering av han som person som klarerer Per eller Pål og ta del fire av det bygget og ta ansvar for det bygget. Det har jeg dratt med meg inn i prøvenemndsarbeidet. Skal man bli og ha selvstendig tankegang som man skal utdanne folk til så må man ha evnen til å tenke og vurdere og planlegge jobber, og i begynnelsen så jeg for meg om dette er en montør jeg kan ha med meg på et anlegg. (F2-P16)

Nemndsmedlemmet viser til koblingen mellom vedkommendes praktiske erfaring som elektriker og deltagelsen i nemndsarbeidet. Nemndsmedlem F3 nevner at lærlingene kan bli spurt om ting på fagprøven hvor de blir konfrontert med sider ved yrkeslivet som de ikke har erfaring med. I forlengelsen av det viser nemndsmedlem F5 til eksempler hvor lærlinger som ble sett på som flinke og som kunne blitt ansatt, allikevel har strøket på fagprøven. Det er også vanlig blant nemndsmedlemmene å stille spørsmål om de kunne bruke denne lærlingen i en elektrikerjobb. Disse tre ovenstående situasjonene er eksempler på hvordan vurdererne ser sammenhenger mellom fagprøven og yrkeslivet. Et siste eksempel handler om kjennskap til firmaet lærlingen jobbet selv om det ikke ble brukt så ofte i vurderingene.

Rater M4 viser hvordan yrket trekkes inn i vurderingene og uttrykker hva som er vesentlig å vurdere «Jeg vil ha en som kan banke kabel og som kan trekke og som kan koble ute jeg. Ikke sitte på et kontor og regne.» (M4-P18). Rater M4 er opptatt av om de valgte løsningene ville fungert, noe som er vanskelig å finne ut basert på en skriftlig besvarelse. Men rater trekker inn yrket og ser besvarelsen i en større sammenheng knyttet til selve utøvelsen. Rater M3 bruker sin erfaring fra yrket i vurderingen gjennom «å visualisere» (tatt opp i kapittel 5.4.1) om prosessen som er beskrevet ville fungere i praksis. Rater gir uttrykk for å prøve å forstå mer enn det som er skrevet ned på papiret: «[...] mere den følelsen av hva han kan og da må jeg på en måte bruke den erfaringen jeg selv sitter med. Jeg prøver å se for meg dette visualisert da rett og slett.» (M3-P18)

### 5.5.2 Diskusjon om vurderernes forutsetninger for helhetlig vurdering

Den påfølgende diskusjon om vurdererens forutsetninger omhandler *fagperson som vurderer*.

I dette avsnittet skal vi se på hvordan vurderernes *faglige bakgrunn* virker inn på en helhetlig vurdering.

Vurdererne i alle de tre vurderingssystemene bygger på sin *faglige bakgrunn* samt hensynet til den konteksten vurderingen skjer i, noe som er tråd med det Wolf (1995) påpeker at vurdererne opererer med holistisk sett av konsepter. Vurderernes erfaringer samt normer etablert gjennom diskusjoner og refleksjoner vurdererne imellom gjør at de i fellesskap bestemmer om lærlingene innehar den aktuelle kompetansen (Crisp, 2012; Deichman-Sørensen et al., 2011; Nyen & Tønder Hagen, 2014). Det er i tråd med det som Wolf (1995) karakteriserer som å være mer aggregerte i sine vurderinger. Rafoss og Witsø (2014) har undersøkt prøvenemnder og viser i sin studie til at det å følge loven i vurderingsarbeidet ikke

praktiseres mer hos prøvenemndsmedlemmer med lang erfaring, de lener seg kanskje snarere til sin faglige bakgrunn og har en større åpenhet for hvordan lærlingene uttrykker seg. Det ser O'Grady og Alwis (2012) i sammenheng med helhetlig vurdering, som en «... professional body» som gjør vurderingene og åpenheten gir rom for å bruke mer av seg selv i helhetlige vurderinger. Jeg har tidligere vist til nemndsmedlem F3s som ser utfordringer i vurderingene når lærlingene ikke klarer å uttrykke seg muntlig, men at førsteinntrykket av praktisk fagutøvelse har stor betydning for den helhetlige vurderingen (Wood, 2014; Nore og Lahn, 2015). Wood mener det er viktig å ha en bevissthet om at feil i vurderingen kan oppstå ved førsteinntrykket. Vurdererne i studien har en positiv innstilling til å gi lærlingene en god opplevelse under fagprøven, men ser det som vanskeligere under en eksamen eller MECVET-test.

Enten vurderernes bakgrunn er fra *skole eller arbeidsliv* så kan Beckett og Hager (2002) vise til at helhetlige vurderinger på arbeidsplassen involverer flere faktorer. Det innebærer også at vurdererne forholder seg til de endringer som skjer i arbeidslivet og sikrer at vurderingene er relevante og meningsfulle. Nemndsmedlem F5 stiller spørsmål om skolefolk kan gjøre den jobben. På samme måte kan det diskuteres om det har noen betydning å ha en blanding av vurderere fra skole og arbeidsliv som kan lære av hverandre og jobbe med helhet versus detaljer ved vurdering av eksamen slik sensor E3 viser til (se også Tsagalidis, 2008). Rater M2 har en opplevelse av å bli snillere i vurderingene på grunn av sin bakgrunn fra skolen. Samtidig er det ofte snakk om at vurdere både må ha kunnskaper i det feltet de skal vurdere og vurderingstekniske ferdigheter (Jones, 1999). Rater M4 var ikke tilknyttet skoleverket og benyttet sin arbeidserfaring som grunnlag. Spørsmålet kan stilles om en utøvende elektriker uten vurderingserfaring kan gjøre like god vurdering av en lærling som en med lang erfaring eller bakgrunn fra skolen og om vurderingene er like helhetlige.

Vurdererne henviser ofte til eksempler fra *yrkeslivet* i vurderingene og opererer med en holistisk modell uten å tenke i atomiserte enkeltelementer (Wolf, 1995). Yrket utgjør en referanse i forhold til den konteksten man er i når vurderingene gjøres og utgjør en støtte for å kunne se og vurdere innholdet i lærlingenes besvarelser. Det innebærer også forskjeller i vurderernes måter å se på enkeltbesvarelsene og medfører en viss subjektivitet i de helhetlige vurderingene, noe O'Grady og Alwis (2012) ser på som en styrke. Samtidig viser Laming (2004) til at det å trekke inn egne erfaringer kan gi grunnlag for feilvurderinger som kan bli forsterket ved repeterende handlinger. Her vil graden av skjønn være fremtredende og bli en del av vurderingsarbeidene (Grimen & Molander, 2008). Også Deichman-Sørensen et al.

(2011) fant at prøvenemndsmedlemmene utøvde skjønn som en del av den helhetlige vurderingen. Om det er svakt eller sterkt skjønn (Dworkin, 1978) avhenger av hvordan den enkelte vurderer utfører sin jobb, men også av påvirkning fra rammene de opererer i. HELVYRD-studien viste at utøvelse av skjønn var fremtredende i alle vurderingssystemene selv med de rammer og føringer som ligger i retningslinjer for arbeidet. Man kan tenke seg at skjønnnet blir svakere ved de mer rigide MECVET-testene, men også der er skjønn en del av vurderingen. Vurdererne i studien peker på de åpne oppgavene som krever at kandidatene viser flere sider ved en yrkeskompetanse og at det igjen krever skjønnbasert vurdering. Når kompetansen også knyttes til en framtidig yrkesutøvelse mener vurdererne at de vurderer mer helhetlig basert både på bruk av kriterier og skjønn. Å etterstrebe vurdering med mindre rom for skjønnsutøvelse vil svekke mulighetene for helhetlig vurdering, og vurdererne gis mindre frihet til å bruke sine erfaringer og kompetanse om yrkeslivet i vurderingsprosessen. Det forutsetter som tidligere nevnt at vurdererne har en oppdatert og relevant kompetanse. Rauners (2013) studie viste at ratere endret sin rating før og etter en ratertrening hvor interrater-reliabiliteten ble høyere. Her kan det stilles spørsmål om skjønnsutøvelsen i vurderingene blir sterkere eller svakere i en slik prosess. Mest sannsynlig ble skjønnsutøvelsen svakere, og demed muligheten til å foreta helhetlige vurderinger. At enigheten økte betyr ikke nødvendigvis at raternes vurderinger er mer reliable. Et siste aspekt er to-sensorsystemet som har en justerende effekt og kan påvirke skjønnnet til den enkelte. Her viser Johnson (2008a) til viktigheten av å gi rom for vurderernes personlige erfaring i vurderingene og å bruke erfaringer fra yrkeslivet. For at tilliten til skjønnsutøvelse kan opprettholdes er det en forpliktelse for vurderere å ha kontakt med yrkeslivet for å holde seg oppdatert om den faglige utviklingen.

Ifølge Biemans et al. (2009), Boud og Hawk (2003) og Shepard (2000) er det av betydning å utvikle vurderingssystemer som er valide og reliable for den vurderingen som skal gjøres. Det er vesentlig at vurdererne har anledning til å trekke inn referanser og erfaringer fra yrkesutøvelsen. Hvis ikke vurdererne føler nok tilhørighet eller eierskap til vurderingsformene kan det hindre den enkelte vurderer i å bruke seg selv som en profesjonell vurderer og dermed vil den helhetlige vurderingen kunne svekkes.

### 5.5.3 Funn knyttet til helhetlig vurdering ved forutsetninger

Fagpersonen som vurderer legger i stor grad opp til helhetlig vurdering, uavhengig av vurderingssystem. I vurderingsprosessen bruker de seg selv og sin faglige bakgrunn





(«professional body»), men hvordan de bruker fagkompetansen påvirkes av vurderingssystemet. Prøvenemndsmedlemmenes nærhet til lærlingene i testsituasjonen gjør at de kan observere både implisitt og eksplisitt uttrykt kompetanse og at lærlingene ivaretas av en erfaren fagperson. Sensorene og raterne forholder seg mer til kompetansemålene og MECVET-kriteriene enn seg selv som fagpersoner i vurderingen.

Vurderernes bakgrunn fra skole eller arbeidsliv hadde betydning for vurderingene. For eksempel tenkte vurderere fra næringslivet mer helhetlig, mens lærere var mer fokusert på detaljer. En rater som både er i skolen og i næringslivet uttrykte at han ble snillere i vurderingen ved å anvende et «skoleperspektiv». Prøvenemndsmedlemmer stilte kritiske spørsmål om «skolefolk» kunne bedømme fagprøver. Samtidig ble det blant andre vurderere sett positivt på å ha en blanding av skolefolk og næringslivsfolk som lærte av hverandre, hvor noen var gode på vurderingsteknisk kompetanse og noen på fagkompetanse. Uavhengig av bakgrunn pekte vurdererne på betydningen av å holde seg faglig oppdatert fordi vurderere bruker eksempler og erfaringer fra yrkeslivet som referanseramme i vurderingene. Det handler om å se helhetlige jobber og sluttkompetanser, det handler om selvstendighet og det handler om hvorvidt lærlingen er klar for å gjøre en jobb. Som tidligere nevnt, visualiserte en rater løsningsforslaget til lærlingen som del av en helhetlig vurdering basert på personlig erfaring, subjektivitet og skjønnsutøvelse. Skjønnsutøvelse var framtrødende og kan bidra til helhetlige vurderinger, men det forutsetter at vurdererne har tett kontakt med yrkeslivet og utnytter de mulighetene som ligger i vurderingssystemene (nevnt under 5.2.4.). Ratertrainingen i MECVET gjorde at raterne fikk mindre rom for skjønnsutøvelse og dermed helhetlige vurderinger.

Oppsummert så vil funn presentert under hvert tema (kapittel 5.1 – 5.5) trekkes inn i en sammenfattende diskusjon i kapittel 6.

## 6. Sammenfattende diskusjon og studiens funn: Helhetlig vurdering i et yrkesdidaktisk perspektiv

I kapittel 5 ble vurderernes konkrete handlinger presentert og diskutert i lys av helhetlige vurdering. Sammenfattende diskusjon av funn vil skje under fire overordnede dimensjoner (6.1-6.4), hvor de yrkesdidaktiske perspektivene redegjort for i kapittel 2.4 (Gessler & Herrera, 2015) trekkes inn. Enkelte nye forskningsbidrag og offentlige dokumenter trekkes inn i diskusjonen. Kapittel 6.5 er presentasjon av sentrale funn. Figur 37 illustrerer sammenhengen.

<b>Funn helhetlig vurdering fra de enkelte tema (Kapittel 5.1-5.5)</b>		<b>Overordnede dimensjoner i en sammenfattende diskusjon. (Kapittel 6.1-6.4)</b>		<b>Yrkesdidaktiske perspektiver (Kapittel 2.4)</b>
Innhold  Rammefaktorer & kontekst  Mål og kriterier  Metoder  Forutsetninger		Ny forskning ----- Faglig innholds mangfold og kompleksitet  Vurderingssystemenes utfordringer ved komplekse yrkesutøvelser  Vurderingspraksiser  Vurderernes profesjonell erfaringsbakgrunn		Læringsutbytte orientering  Historiske og kulturelle karakterer  Forming og utvikling av yrker (horisontale)  Faglige nivåer (vertikale)  Ulike opplæringsfaser (temporære)  Arbeidslivets naturlige endringer
Studiens funn: helhetlig vurdering i et yrkesdidaktiske perspektiv (kapittel 6.5)				

Figur 37. Skissen viser de fire overordnede dimensjonene i den sammenfattende diskusjon, med utgangspunkt i funn om vurderernes helhetlige vurderinger fra tema med tilhørende kategorier samt de yrkesdidaktiske perspektiver. Avrundes med en syntetisering av studiens funn og en konkludering opp i mot forskningsspørsmålene.

## 6.1 Mangfold og kompleksitet ved faglig innhold: Elektrikernes yrkeskompetanse

Det første funn viste mangfold og kompleksitet i faglige innholdskomponenter som fremkom i vurderernes vurderinger i alle tre vurderingssystemene. Det stemmer med redegjørelsen for det komplekse kompetansebegrepet (Kunnskapsdepartementet, 2016, s. 28; NOU 2018:2, 2018, s. 14; NOU 2019:2, 2019, s. 17). Igjen understøtter det behovet for den tette forbindelsen til arbeidslivet. En identifisering av faglig innhold kan være med på å legitimere hva som er relevant i elektrikerfaget, noe som understøttes i perspektivet om (1) orientering av læringsutbytte (Gessler & Herrera, 2015, s. 153). Men mangfold og kompleksitet i innholdskomponenter utfordrer vurdererne i hvordan de skal forholde seg til elektrikernes yrkeskompetanse. Spørsmålet kan stilles om det går an å vurdere så mange innholdskomponenter helhetlig, og om noen av komponentene vil være mer egnet for helhetlig vurdering enn andre. I den sammenheng vil pågående diskusjoner blant vurdererne om faglig innhold være av betydning for i størst mulig grad å kunne se mer helhetlig på vurderingene, og for å sikre at arbeidslivets krav blir tilfredsstillt. Samtidig gir innholdskomponentenes gjensidighet behov for avveininger i den aktuelle vurderingskonteksten, noe som har likheter med det Mulder et al. (2007, s. 82) trekker frem som den ubalanserte vektleggingen som skjer i vurdering av kompetanse. Det gir rom for å legge et «tyngdepunkt» inn i vurderingen, slik sensor E2 påpeker. Men siden det fremkom at den overordnede faglige innholdskomponenten i elektrikerfaget er el-sikkerhet (som må ses i lys av de lover og forskrifter som elektrikerfaget styres av), vil det prege vurderingene sett i lys av kriterier og skjønnsutøvelse. Dermed vil den øvrige gjensidigheten mellom innholdskomponentene ikke gjøre det mulig å lage en ferdig skrevet innholdsliste, da det i stor grad vil være kontekstavhengig. Med yrkeskompetansens kompleksitet vil behovet for at vurdererne tilnærmer seg vurderingen på en mer pragmatisk måte være til stede. Hager (2017, s. 209) underbygger dette ved å vise til behovet for et integrert syn på kompetanse. Det innebærer blant annet å ha en situasjonell forståelse («situational understanding») i ulike kontekster. Et helhetlig integrert syn på kompetanse vil være en motvekt til en atomisert og rigid vinkling som fører til en fragmentering av elektrikerfagets yrkeskompetanse. Hager, for eksempel mener det er vesentlig å få samkjørt alt («putting it all together») (s. 209) for å ivareta flest mulige sentrale nøkkel-kompetanser. Johnson (2008, s. 29) uttaler at helhetlig vurdering innebærer å få alt «[...] into a final judgment». Lum (2013, s. 1202) peker på at det integrerte konseptet om kompetanse innebærer å akseptere dens kompleksitet ved vurderingene. Samtidig påpekes betydningen, som flere andre forfattere har uttalt, av at det

ikke er mulig å utvikle helt komplette beskrivelser av en kompetanse\_ (Hager, 2017, s. 212). Det igjen kan utfordre de studiene (Petersen og Rauner: Lindberg) Rauner et al. (2013, s. 23) viste til sitt arbeid med forsøk på å definere sentrale hovedområder innen profesjonelt arbeid.

Det at mangfoldet og kompleksiteten i det faglige innholdet i elektrofaget gjør det utfordrende å utforme en komplett innholdsliste, fører oss også over til det andre funnet. Ved å forholde seg til en kompleks yrkeskompetanse hadde vurdererne en åpenhet i vurderingen av lærlingens besvarelse. Lærlingen kunne presentere løsninger som vurdererne ikke hadde tenkt over, noe som ble tatt hensyn til. Samtidig må det ses i sammenheng med prøvenes representativitet og de kontekstuelle forholdene. Gjennom åpenheten fremkom et mer differensiert syn på den enkelte lærling, noe som i større grad kan sies å ivareta den individuelle kompetansen og den enkelte lærling sin læringsbane (Erauts, 2004, s 804). Åpenheten ovenfor lærlingenes besvarelser kan knyttes til det å etterstrebe en identifisering av yrkesoppgaver i arbeidslivet og lærlingens kyndighetsutvikling (Inglar, 2009, s. 84). Det vil understøtte Gessler og Herrera (2015, s. 153) sitt fokus på at den lærende (1) aktivt skal være med å utforme sitt arbeid og arbeidsplass, noe som er vesentlig i utvikling av lærlingens selvstendighet.

Mangfoldighet og åpenhet i det faglige innholdet gir vurdererne et handlingsrom i vurderingene. Til en viss grad lå det en spenning om vurderernes ulike syn på vesentlige innholdskomponenter for en elektriker, noe rater M3 viste til (ubalanserte vektlegging, Mulder (2007)). Betydningen av kollegiale diskusjoner og organiseringen av to-sensorsystemer var sentralt for bevisstgjøring og samstemthet. Samtidig vil det være viktig i diskusjonene, å være bevisst at fremtidens faglige innhold og utviklingen av elektrikerfaget er i en rivende utvikling, noe som ble tatt opp innledningsvis i kapittel 1.4. Her kan blant annet grenseoppgangen mellom fagområdene innen elektro være et element, noe som kan knyttes til Gessler og Herrera (2015, s. 156) om (3) formingen og utvikling av yrkene. I tillegg vil det ved yrkesdidaktiske tilnærminger være behov for å nyansere mellom ulike yrker, samtidig som det er vesentlig å se yrket i en større sammenheng med andre yrkesgrupper. Det vil ha direkte betydning for vurderere som må ta hensyn til den stadig endrede arbeidsvirkelighet som møter en elektriker. Det innbefatter en økt globalisering og samarbeid på tvers av ulike land, noe som kan resultere i at man tilnærmer seg hverandre og får en økt felles forståelse. Det underbygger betydningen av Gessler og Herrera (2015, s. 156) sitt perspektiv om (6) arbeidslivets naturlige endringer. Teknologi-drevne trender (Nærings- og fiskeridepartementet, 2017, s. 65) med revolusjon innenfor produksjon, hvor blant annet

automatisering og digitalisering vil påvirke både produkter og verdikjeder, får betydning for utviklingen av elektrikerfaget. Digitaliseringen handler ikke bare om å bruke nye teknologier, men like mye om viljen og evnen til å gjennomføre endringene for å forbedre og fornye (*Digital21*, 2018, s. 9). Videre uttrykkes det behov for dyktige fagarbeidere, som utover de rent faglige kvalifikasjoner, også har « [...] kompetanse som omfatter helhetsforståelse, innovasjonsforståelse, kvalitets- og effektivitetsforståelse, ansvars- og ledelseskompetanse og dokumentasjonsforståelse.» (Nærings- og fiskeridepartementet, 2017, s. 132). Det pekes også på betydningen av planlegging, selvstendighet og det å beherske flere teknologier, i tillegg til en rekke andre beskrivelser av en kompetent industriarbeider (Nærings- og fiskeridepartementet, 2017, s. 132).

Samlet sett viser styringsdokumentene en sammensatt kompetanse for fremtidens fagarbeidere, noe som vil ha en direkte innvirkning på vurderernes praksis i vurderingene, og utviklingen av vurderingssystemer som i størst mulig grad skal sikre en representativitet. Perspektivet om (2) kulturell og historisk karakter ved yrkesdidaktikken (Gessler & Herrera, 2015, s. 153) viser at den er prisdelt den pågående utviklingen. På bakgrunn av diskusjonen i dette kapitlet er det mulig å identifisere noen sentrale innholdselementer som inngår i en helhetlig vurdering. Med store forandringer i elektrikerfaget vil en kontinuerlig oppdatering av forståelsen av hvordan man skal utvikle fag- og yrkesopplæring være sentralt for å møte fremtidens krav og kompetansebehov om faglig innhold. En tett forbindelse til arbeidslivet for identifisering av yrkesoppgaver og faglig innhold for en elektriker vil være av betydning for å fremme kyndighetsutvikling. Det må skje i sammenheng med behovet for en åpenhet og tilpasning til ulike kontekster sett i lys av yrkeskompetansens kompleksitet og mangfoldighet. Vurderernes helhetlige vurderinger må tilpasses den faglige utviklingen. Vurderingssystemenes representativitet vil være en påvirkende faktor, noe som skal diskuteres ytterligere i neste kapittel.

## 6.2 Vurderingssystemenes utfordringer ved en kompleks yrkeskompetanse

De tre vurderingssystemene viste at vurderingsformer og oppgaveutforminger er under en stadig diskusjon og evaluering. Det kan forstås slik at det har en sammenheng med at vurdererne gjennom sin faglig kompetanse innen elektrikerfaget ser endringene i arbeidslivet. Som diskutert i kapittel 6.1, vil teknologiutvikling, endrede arbeidsformer og mer tverrfaglighet innebære behov for økt kunnskap og endrede mønstre i arbeidsutførelsen. Endringene i arbeidsmarkedet må gjenspeiles i de anvendte vurderingssystemene, slik at både lærlinger og vurderere får operere innenfor rammer som oppleves meningsfulle. Van der

Vleuten, Sluijsmans og Joosten-ten Brinke (2017, s. 608) viser til endringer hvor det legges vekt på autentiske vurderingsformer i ulike kontekster for å ivareta mangfoldet i en kompetanse. Gipps (1999, s. 384) viser også til et bredere perspektiv på vurdering og fremmer blant annet det mer sosiokulturelle synet på vurdering. Men en perfekt vurderingsform kan ikke utformes (Eraut, 2004, s. 803). Det trengs et mangfold av vurderingssystemer (multiple vurderingssystemer) som gir et bredere spekter av vurdering (Shepard, 2000, s. 8). Samtidig vil det være nødvendig å stille spørsmålet om det enkelte vurderingssystem kan ivareta lærlingenes yrkeskompetanse og i hvilken grad. Er for eksempel fagprøven på en prøvestasjon godt nok? Er dagens form på fagprøven og eksamen god nok til å gjøre helhetlige vurderinger? Får man vurdert aktuell yrkeskompetanse for nåværende og fremtidige yrker? I hvilken grad kan skriftlige tester som MECVET avdekke en kompleks yrkeskompetanse? Formålet til de tre valgte vurderingssystemene i denne studien vil være ulike: Eksamen skal dokumentere kunnskaper, MECVET er mer diagnostisk og kan knyttes til det mer formative og fokus på videre læring og fagprøven er en sertifisering for arbeidslivet. De kan ses på som komplementære og dermed sammen fungere som en helhet i vurderinger. Det ble eksemplifisert ved sensor E2, som uttrykte at skriftlige eksamener ikke alene kunne gi nok grunnlag for å vurdere en lærling. Representativiteten til det enkelte vurderingssystem kan i stor grad knyttes til om vurdererne opplevde å gjøre vurderingene av lærlingene i tett tilknytning til elektrikernes reelle yrkesutøvelse. Vurdererne viste også evne til å tilpasse og optimalisere innenfor de enkelte vurderingssystemene for i størst mulig grad skape meningsfulle og gode vurderingssituasjoner, noe som kom frem spesielt ved fagprøven. Det skjedde uten at det gikk utover lojaliteten til det enkelte vurderingssystem. Samtidig, i et yrkesdidaktisk perspektiv, vil formen på vurderingene i elektrikerfaget være av vesentlig betydning for å kunne følge den (6) nasjonale og internasjonale utviklingen innenfor yrket (Gessler & Herrera, 2015, s. 157). Det kan bety at det ikke må være for store avstander mellom oppgaveutviklere, vurderere og det reelle elektrikerfaget, hvor van der Vleuten et al. (2017, s. 625) snakker om å ha en «intense collaboration» mellom de ulike partene i opplæringen. Mulder (2007, s. 82) peker på at vurdering ofte skjer i simulerte prosesser og garantiene for at lærlingene vil være kompetente i en virkelig elektrikerjobb er ikke absolutte. Det kan underbygges av Olsen (2009, s. 7), som uttrykker at en fullverdig vurdering kun kan skje der hvor yrkesutøvelsen skjer. Disse synspunktene kan knyttes til alle tre vurderingssystemene. Samtidig viser Palm (2008, s. 9) til behovet for å tydeliggjøre hva som ligger i begrepene autentiske vurderingssituasjoner og prestasjonsbasert vurdering, da ikke helt klare definisjoner og forståelser ligger til grunn i alle tilfeller. Diskusjonen viser

kompleksiteten ved å utvikle og gjennomføre vurderinger i ulike vurderingssystemer. Som det kommer frem av min studie vil alle vurderingssystemene kunne fungere i en helhetlig vurdering, men det vil være i ulik grad hvor ulike påvirkninger spiller inn.

Funn viste til dels en sammenheng mellom resultatene i de anvendte vurderingssystemene, blant annet mellom eksamen og fagprøven. Det kan uttrykke noe om vurderingssystemenes representativitet. Hvis ikke avvikene er for store og det fremstår et mønster hvor lærlingen skårer nogenlunde likt på de ulike vurderingene, kan det være et uttrykk for at vurderingssystemene treffer bra (god sammenheng). Samtidig kan det knyttes til Gessler og Herrera (2015, s. 156) sitt yrkesdidaktiske perspektiv på betydningen av (3) transformasjon av arbeidssituasjoner fra elektrikerfaget som anvendes i en vurderingskontekst for både vurderere og lærlinger. Sammenhengen mellom vurderingssystemene kan også utvikles gjennom at summative og formative vurderinger i større grad kan integreres og virker i en helhet, noe Throndsen et al. (2009) påpekte betydningen av. Det kan øke nærheten til elektrikerfaget og ikke minst kontinuiteten i opplæringen. Med bakgrunn i synspunkter om begrepet yrkesdidaktikk og uttalt behov for ulike vurderingssystemer, vil det ved en større grad av automatisering og mekanisering av vurderinger være mer utfordrende å gjøre en reell vurdering av en aktuell sammensatt yrkeskompetanse. Det kan argumenteres for at del-prøver og systematisk tilbakemelding på oppgaveløsning kan ha en formativ effekt, og på sikt understøtte summative prestasjoner. Men det er vesentlig å være klar over de begrensninger det kan ha med en mer teknisk tilnærming til vurdering og spesielt ved de summative vurderingene. Elektrikerfagets utvikling, og dermed yrkesdidaktikkens utvikling, gir behov for stadig tilpasninger av vurderingens former som sikrer representativitet og rettferdighet. Det for å sikre at en fremtidsrettet vurdering av en kompleks yrkeskompetanse finner sted. Det vil igjen sikre at det kan være en sammenheng mellom resultatene i de enkelte vurderingssystemene, selv om ulike sider av en yrkeskompetanse blir vurdert i de ulike vurderingssystemene.

Studien viste til dels et kritisk blikk hos vurdererne på alle typer utforming av oppgaver. Men oppgaver med casebaserte utforminger fra en elektrikers arbeidshverdag ga vurdererne en større opplevelse av å være nær yrkesutøvelsen (mer autentiske). Den skriftlige eksamen formet som en arbeidsordre var godt likt, og den var lett å kjenne igjen fra yrkeslivet for både lærlinger? og sensorer. Det ga sensorene anledning til å gjøre vurderingen i tett tilknytning til egen yrkesutøvelse. Den casebaserte formen til MECVET ble i større grad kritisert på grunn av mengde tekst i beskrivelsen. Ved å prøve å dekke alle sider av et arbeidsoppdrag for en

elektriker i en gitt kontekst med en utfyllende tekst, kunne det i større grad virke mot sin hensikt og være mer hemmende for lærlinger og vurderere. Utfordringen kan ses i lys av Wolf (1995, s. 55) og Hager (i Jones, 1999, s. 154) som peker på den umulighet det er å beskrive et arbeidsdomene fullt ut. Oppgavene anvendt under fagprøven ble kritisert, noe som også hang sammen med anvendelse av prøvestasjoner, men de fungerte som et utgangspunkt for en vurderingssituasjon. Med prøvenemndsmedlemmenes anledning til direkte kommunikasjonen med lærlingen, bidro det til å skape en best mulig situasjon for vurdering av lærlingene. I større grad kunne de være pragmatiske, hvor oppgavene fungerte som en katalysator for å vurdere en større helhet av læringen ved ha en muntlig gjennomgang. Ved den stadige utviklingen av elektrikerfaget må den yrkesdidaktiske oppmerksomheten omkring oppgaver ta hensyn til kontekstene vurderingene skjer i og samtidig sikre en best mulig (3) nærhet til yrket (Gessler & Herrera, 2015, s. 156). De som utvikler oppgavene, vil ha et ansvar for å sørge for at ingen av lærlingene kommer i vurderingssituasjoner som kan virke fremmedgjørende, men må være «suited to the task» (Dobson, 2010, s. 35). For å følge opp Dobson viste min studie at vurdererne etterstrebe vurderingssituasjoner som støtter opp under lærlingenes læreprosesser. Det krever at alle oppgaver har nok og riktig informasjon, hvor hjelpemidler og tid også må tas hensyn til.

Alle vurderingssystemer har et skriftlig element (se 5.1.1 og 5.2.1). Funn i studien viste at det var spesielt to former for skriftlighet som var tilstede: (1) dokumentasjonsskriving og (2) beskrivelser og redegjørelser av faglige begrunnelser i besvarelser av oppgaver. Vurdererne i denne studien uttrykte at skriftlighet i lærlingenes besvarelser påvirker dem. De påpekte også at det legges vekt på mestring av skriving i yrket. Først og fremst gjelder det dokumentasjonen. Skriftlighet tilknyttet beskrivelser og redegjørelser i besvarelsene vektlegges annerledes. Spesielt fremkommer det når vurdererne mener å kunne lese mellom linjene, hvor det skjer en visualisering omkring lærlingens løsninger i prosessen med å vurdere besvarelsene. Det betyr at den skrevne teksten ikke står alene i vurderingene, men heller bidrar til å sette vurderer inn i en kontekst hvor skjønn og tolkninger er en del av vurderingen. Flere vurderere ga uttrykk for at en muntlig tilleggsdel ville bidratt positivt for lærlingen, noe som uttrykker at bare skriftlig testing kan virke begrensende på å få frem lærlingenes yrkeskompetanse. Så begge former for skriftlighet påvirker vurderernes helhetlige vurderinger. Hvilke former for vurdering som kan bygge oppunder betydningen av skriftlighet for elektrikerne kom ikke tydelig frem. Men med bruk av arbeidsordre som et utgangspunkt for en vurderingsform ga sensor E3 uttrykk for at det skriftlige hadde hevet seg. Studien til



Hellne-Halvorsen, Lahn og Nore (In press) viser forskjellige skrivekulturer blant ulike yrker hvoriblant elektrikeryrket. Studien viser viktigheten av å ha kombinasjonen av den generelle og profesjonelle skrivekompetansen, kalt den hybride skrivekompetansen, i et moderne arbeidsliv. Det har paralleller til funn fra vurderernes uttalelser om viktigheten av at elektrikerne kan uttrykke seg skriftlig.

Ved å ha vurderingssystemer med åpne oppgaver får man rom for større mangfold i lærlingenes besvarelser. Det påvirker og utfordrer vurdererne i større grad i deres vurderingsarbeid. Ulike oppfatninger og tolkninger rundt både oppgaver og besvarelser oppstår. Samtidig vil det ved en slik åpenhet kunne gi bedre muligheter for å følge utviklingen i elektrikerfaget og lettere innpasse nye situasjoner. Det er relevant til Gessler og Herrera (2015, s. 156 og 157) sine perspektiver om både (6) naturlige endringer i arbeidslivet og (3) yrkenes utvikling, hvor innovasjoner og utvikling av produkter er sentralt. Behov for mangfold av vurderingssystemer er belyst, og Baartman et al. (2007) sine kriterier for utvikling av gode vurderingsformer (kapittel 3.1.5) kan utgjøre et utgangspunkt for utvikling av alle former for vurdering som er knyttet til kompetanse.

Det er mange utfordringer for vurderingssystemene som er utformet for å vurdere en kompleks yrkeskompetanse. Denne studien viser muligheter og begrensninger med alle tre. I en helhetlig vurdering vil fagprøven i utgangspunktet kunne favne mest for å vurdere en yrkeskompetanse, men begrenser seg både gjennom å anvende prøvestasjon og forme på oppgaver. Eksamen får vurdert kunnskaper og har en åpenhet som kan gi rom for helhetlige vurderinger, men det begrenser seg ved at den praktiske faktoren er borte. MECVET sin form gir større begrensninger for å tenke helhetlig vurdering både gjennom oppgavene og strukturen for vurdering.

### 6.3 Vurderingspraksiser

Vurdererne hadde en bred og åpen tilnærming i gjennomføringen av vurderingene. Men ulike strategier i vurderingene, som for eksempel prøvenemndsmedlemmens tilnærming til planleggingsdel, ble benyttet for i størst mulig grad ivareta en elektrikers yrkeskompetanse innenfor rammene til de tre vurderingssystemene. Det bygger oppunder Ecclestone (2010, s. 339) og Biemans (2009, s. 269) sine oppfatninger om behovet for å utvikle vurderinger og vurderingskontekster som må tilpasses den komplekse yrkeshverdagen.

Et funn ved planlegging var vurderernes behov for å sette seg godt inn i oppgaveteksten med tilhørende kriterier og lærlingenes besvarelser. Oppmerksomheten var rettet inn mot det faglige innholdet. Samtidig innebar planleggingen å ha en bevissthet om muligheter og begrensninger som lå i det aktuelle vurderingssystemet. Som et eksempel fikk den generiske (fagovergripende) kompetansen større plass på fagprøven enn i de to andre vurderingssystemene. Vurdererne jobbet med å se flest mulig sider av besvarelsen, noe som innebar både deloppgave-tenkning og helhetlig tenkning. Men de fleste strategier handlet om å ha en samlet vurdering på besvarelsen til slutt. Strategien til raterene skilte seg mest ut da formen på MECVET-testen kunne virke mer styrende for hvordan planleggingen kunne gjennomføres grunnet den konkrete fremgangsmåten vurderingsskjemaene var utviklet på. Så vurderernes anvendte strategier kan ses i lys av Gessler og Herrera (2015, s. 153 og 156) sine perspektiver om både (1) læringsutbytte-orientering og (4) faglige nivåer. Det vil si at vurdering av faglig innhold innebærer behovet for å ta gjeldende vurderingskontekst med aktuelt vurderingssystem i betraktning når man planlegger.

Et annet funn er utformingen av kompetansemål og MECVET-kriteriene på den ene side og påvirkningen det har på vurderernes vurderingspraksis på den andre. Det er forskjeller mellom de norske vurderingssystemene og det tyske, men det ble påpekt en viss sammenheng. Selv om kompetansemålene var kjent blant vurdererne ble de ikke anvendt direkte i vurderingen ved eksamen og fagprøven. MECVET-kriteriene og kriteriespesifikasjonene, derimot, la i større grad føringer på raternes praksiser. Ved begge varianter ble det stilt spørsmålsteget ved om graden av vanskelighet ved noen av de kunne være for høy i forhold til forventet yrkeskompetanse for en elektriker. På den ene siden kan kompetansemålene ses på som en overordnet ramme som gir rom for tolkninger og meninger om elektrikernes yrkeskompetanse. På den andre siden kan antall kompetansemål virke reduksjonistisk og for løselige og det kan oppleves for omfattende til å se helhetlig på det faglige innhold. Mulder (2007, 82) uttrykker at formuleringene av kompetansemål kan bli for generelle og påvirke tydeligheten i vurderingene. Spørsmålet kan stilles om utformingen av kompetansemål i større grad kan tilnærme seg utformingen av MECVET-kriterier med tilhørende kriteriespesifikasjoner og vice versa. Men rom for tolkninger og tilpasninger i ulike vurderingskontekster må ivaretas, noe som fremkom i studien og som flere forfattere påpeker viktigheten av. Det handler om en balansegang hvor utforming av en veiledende ramme ikke blir for styrende, slik MECVET-kriteriene til dels representerer, og ikke for løst formet, slik kompetansemålene kan være en representant for. Det handler om å utvikle en form som kan gi

vurdererne nok rom til å finne det faglige «tyngdepunktet» i den til enhver tid gjeldende vurderingskonteksten. Tigelaar og van der Vleuten (2014, s. 1262) bekrefter dette ved å påpeke nytten av å ivareta vurderernes egen innsikt i yrket. Utviklingen av en ramme for vurdering må ivareta rommet for ulike perspektiver, noe som også gir vurdererne anledning til å se mer helhetlig på besvarelser. Samtidig har konteksten betydning for hvordan vurderer anvender kompetansemålene i vurderingssituasjonene. Sensor E1 ga uttrykk for forskjeller når du ser og ikke ser lærlingen. Det kan knyttes til Gessler og Herrera (2015, s. 153) sitt perspektiv om (1) læringsutbytte-orientering og bevisstheten om faglig innhold som framkommer ved eksamen og ved fagprøven.

Funn om utforming og anvendelse av kriterier i vurderingspraksisene viser at de ikke kan bli for konkrete eller for stramme, skal vurdering ivareta en kompleks yrkeskompetanse, noe som i større grad samsvarer med den presenterte forskningen. Ved en større grad av stramhet i kriteriene (atomiserende eller reduksjonistiske) påvirker det fleksibiliteten og tilpasningene i vurderingsprosessen. Hager og Gonczi (1996, punkt 2) støtter til en viss grad atomisering, men bare så lenge helheten ivaretas til slutt. Man kan argumentere for at tydelige kriterier også kan gi verdifull informasjon til vurdererne, men hvis det i for stor grad er bestemmende for hvordan de skal anvendes, kan det gå utover vurderers egen autonomi. Funnene i denne studien viser et mangfold i utforming og anvendelse av kriterier i de tre vurderingssystemene. Pendelmetaforen (Dobson, 2010, s. 33) var et forsøk på å illustrere hovedtendensen i de tre vurderingssystemene (se figur 37); den holistiske (fagprøven) på den ene siden og den mer analytiske (MECVET) på den andre siden. Eksamen kan sies å være et sted midt imellom. Det fremkom forskjeller i anvendelsen internt i vurderingssystemene også. Men det tette samarbeidet mellom vurderingskolleger ved eksamen skapte et eieforhold og lokal tilhørighet til anvendte kriterier i aktuell kontekst, noe som skapte en enighet. De forhåndsskrevne kriteriespesifikasjoner ved MECVET påvirket vurderernes bevissthet om faglig fokus, og anvendelsen av dem var i stor grad gitt. I mindre grad fremkom det et eieforhold og i større grad en form for uniformering og mekanisering. Graden av selvstendighet i egen vurderingspraksis var mindre enn ved eksamen. Fagprøvens kriterier som nemndsmedlemmene har utviklet og forhåndsskrevet, har paralleller til MECVET, men anvendelsen har mer paralleller til eksamen. Nemndsmedlemmenes vurderingspraksis er mer pragmatisk i anvendelse av kriteriene og tilpasses den aktuelle vurderingskonteksten. Dermed utvikles det i større grad et eieforhold og lokal tilhørighet til kriteriene. Funnene viser tydelig forskjell i utforming og anvendelse av kriterier i de tre vurderingssystemene. Det gir tre ulike

innfallsvinkler, noe som preger muligheten for å gjøre helhetlige vurderinger. I den presenterte teorien og forskning om anvendelse av kriterier, så vil min argumentasjon være at sensorenes tilnærming ved eksamen ligger nærmest det å være helhetlig i sin anvendelse i vurderingene. Det gjelder både den fortløpende utviklingen og anvendelsen av kriterier som skjer i et vurderingsfellesskap. Det kan også knyttes til Gessler og Herrera (2015, s. 153 og 156) sine forutsetninger om (1) orientering mot læringsutbytter og (3) forming av yrkene.

Det er viktig å ivareta og anvende vurderernes egen erfaring i utviklingen av kriterier. Men funnene viser at det bør skje i diskusjonsfellesskaper med kolleger, og som poengtert i kapittel 6.1 om yrkets utvikling, må ikke diskusjonene virke for uniformerende eller mekaniserende. Vurdererne har behov for fleksibilitet med mulighet for å tilpasse sine vurderinger til (3) endringer og utvikling i elektrikerfaget (Gessler & Herrera, 2015, s. 156). Det innbefatter også en åpenhet for faglige innspill fra lærlingenes besvareelser. Lum (2013, s. 1203) viser til at det er viktig å ha en «expansive mode» hvor vurdereres «facility to select and assign significance to any evidence that is available» (Lum 2013, s. 1203) uttrykker vurderernes rolle og betydning i komplekse vurderinger av en yrkeskompetanse, noe som samsvarer med Wolf (1995, s. 67). Funn viser at vurderingsprosessen derfor må ivareta lærlingenes yrkeskompetanse basert på den praksis vedkommende har hatt i sin opplæring. Behovet for anvendelse av kriterier i helhetlig vurdering så viser studien at det må fordres nok rom, slik at ikke viktige innholdselementer blir utelukket. Det handler om betydningen av vurderernes eieforhold til kriteriene når vurderingen skal ivareta lærlingenes yrkeskompetanse, samtidig som det må være rom for mer den uformelle siden (Colley og Jarvis, 2007, s. 312). For å ha en «expansive mode» vil Eisners (1991, s. 63) *connisseurship* om å verdsette (*appreciation*) mangfoldet være en vesentlig del i det å ha helhetlige vurderingsprosesser. I lys av kriteriediskusjonene må vurdererne ha det rom for å anvende egen fagkompetanse og standarder som en del av vurderingsrammen, noe Wolf (1995, 67) understreket. Funnene viser at vurdererens vurdering påvirkes av samspillet mellom de utviklede offisielle kriterier (som igjen er påvirket av kompetansemål og MECVET-kriterier) og personens egen fagkompetanse. Forskningen til Nore og Lahn (2015, s. 164) bekrefter dette i tilfellet med fagprøven. Min studie utdyper dette ytterligere, og samtidig viser den andre nyanser av samspillet gjennom både eksamen og MECVET-testen. Det blir en pragmatisk tilnærming til den aktuelle vurderingskonteksten, hvor helheten etterstrebes, men påvirkningen fra vurderingssystemene gir ulike rom og muligheter. Mer utdypende om vurderernes egen fagkompetanse og erfaringsbakgrunn anvendt i vurderingene kommer i kapittel 6.4.

Som en del av funnene knyttet til mål og kriterier ble grenseoppgangene mellom *ulike nivåer* i faget med på å prege vurderingspraksisene. Gessler og Herrera (2015, s. 156) sitt yrkesdidaktisk perspektiv om (4) faglige nivåer berører dette, samt betydningen av å ha rett kompetanse til rett jobb. Sammenligning mellom nivåene, som for eksempel fagarbeider og tekniker/saksbehandler/ingeniør kan illustrere kompleksitet i en elektrikers yrkeskompetanse, noe som utfordrer grenseoppganger. Det henger sammen med Mulder (2007, s. 82) sitt argument om svakheter ved analysen av kompetanseprofiler. Eksemplet fra rater M4 og rater M1 om beregninger viser ulik oppfattelse av hva som er nødvendig å ha for en elektriker. Det vil påvirke både vurderernes utforming og anvendelse av mål og kriterier. Funnene viser at det er mest i de skriftlige delene av besvarelsene diskusjonen oppstår om grenseganger og faglig innhold. Det yrkesdidaktiske perspektivet (Gessler & Herrera, 2015, s. 156) som omhandler (3) forming av yrker, viser at lærlingen må bli vurdert i noe som er gjenkjennelig fra egen praksis som elektriker. Vurdererne må trekke inn og vektlegge at rett yrkeskompetanse blir vurdert, noe som viser betydningen av å ha et helhetlig blikk på lærlingenes utførelse, noe som kan knyttes til transfer (Aarkrog 2010, s. 19). Funn om nivåer viser at det ikke helt tydelig kommer frem hvilke grenseoppganger som gjelder. Med den stadige utviklingen innenfor elektrikerfaget vil det være en utfordring å få helt klare grenseoppganger, og dette vil derfor sannsynligvis være til gjenstand for en kontinuerlig diskusjon.

Som jeg var innom i utviklingen av kriterier så viser mitt materiale at samarbeid er viktig fordi vurderingspraksisen ikke er et individuelt ansvar. Det kan underbygges i lys av Mulder (2007, s. 82) som påpeker den utfordring som ligger i det å avgjøre om en kompetanse er nådd eller ikke. Funnene viser at samarbeid mellom vurdererne er høyst til stede i større deler av vurderingsprosessen. Med åpne oppgaver og ulike utforminger for utvikling og anvendelse av kriterier, viser funnene at det i ulik grad etterstrebes en kultur med meningsutvekslinger og tolkningsfellesskap for vurdererne. Et rom for samarbeid mellom vurdererne innebærer et større eieforhold til vurderingsprosessen. Sentralt er to-sensorsystemet. Vurdererne gjør sammenligninger av besvarelser både opp imot sine egne vurderinger, men også mot medvurderer. Det bidrar med nye og andre perspektiver i de enkelte besvarelsene, og er berikende og bevisstgjørende for vurderernes vurderinger. Funnene viser at gjennom samarbeid framkommer også en stor grad av *samstemthet* mellom vurdererne. Der innvirker også vurderernes lange vurderingspraksis. Samtidig vil åpenheten i oppgavene også få frem ulikt syn på faglig innhold (diskutert i 6.1). Grimen og Molander (2008, s. 192) sier det alltid

vil være en grad av uenighet, men den må minimaliseres mest mulig gjennom kommunikasjon. To-sensorsystemet skaper diskusjoner og bevisstgjør vurdererne, noe som gir økt forståelse omkring oppgave, kriterier og besvarelser. Dermed vil det faglige «tyngdepunktet» i en aktuell kontekst lettere kunne bli bestemt. Enkelte funn viste at resultatene hadde vært annerledes uten samarbeid med andre vurderere. De har en justerende effekt på hverandres vurderingspraksis. Det underbygger Gessler og Herrera (2015, s. 153) sitt perspektiv om mangfoldighet og de krav som ligger i (1) orienteringen om læringsutbytter. Samtidig har vurderernes vurderingspraksis et perspektiv om vurderenes egen autonomi i kraft av deres egen fagkompetanse. I den forbindelse kan det stilles spørsmål ved drillingen av raterne i MECVET-testen. Den form for samstemthet er i mye større grad preget av MECVET-testens rammer. Utformingen kan se ut til å utfordre raterernes rom for å gjøre helhetlige vurderinger. Det vil også utfordre vurderernes eieforhold til både oppgave og kriterier, spesielt hvis man opplever at deler av en besvarelse ikke blir ivaretatt.

Med stor grad av samstemhet mellom vurdererne gir det seg utslag i karaktergivningen. Men karaktersystemenes ulikhet gjennom antall graderinger har betydning for vurderernes vurderinger. De seks karaktergraderingene ved eksamen utfordret vurdererne til gå grundigere og mer detaljert inn i besvarelsene for å kunne bestemme en grad. Funn viste at gjennom anvendelsen av arbeidsordreoppgaver medførte det en større anvendelse av hele karakterskalaen. I motsatt retning har fagprøven færre graderinger. Det gir mindre føringer, og karakteren «bestått» virket mer som en sekkepost. MECVET-systemets mekaniserende karaktergiving med avkrysninger og beregninger formet i stor grad vurderernes praksis. Karakteren skal i størst mulig grad speile den helhetlige yrkeskompetanse en lærling skal ha. Spørsmålet er om man det er mulig å lage karaktersystemer som kan fange opp den kompleksiteten som ligger i det å være en utøvende elektriker og gjenspeile den reelle kompetansen, noe Allen (2005, s. 220) påpekte viktigheten av. Gillis og Griffin (2014, s. 62 og 66) viser til generisk og spesifikt grunnlag for karaktersetning og behovet for en bevissthet om kriteriers ulike betydninger når de skal anvendes inn i en karaktersetning. Det vil omhandle Gessler og Herrera (2015, s. 153) sitt yrkesdidaktiske perspektiv om (1) læringsutbytte orientering. Men det framkom ikke noe svar i empirien eller litteraturen på om færre eller flere graderinger på karakterskalaen har betydning for om vurderingen blir mer helhetlig eller ikke. Men det er ingen tvil om at fleksibilitet og tilpasning av karakterer er viktig i vurderingssammenheng. Med vidt forskjellige karaktergraderinger vil det ved alle tre være behov for utvikling slik at man kan tilpasse seg en virkelighet i yrkeslivet.

Vurderernes praksis påvirkes også av opplæringens ulike *faser*. I HELVYRD-studien var det en viss tidsforskjell mellom gjennomføring av de ulike vurderingssystemene. Funn viste at det var en del av strategien å ta hensyn til lærlingenes opplæringsløp i vurderingene. I et yrkesdidaktisk perspektiv om åpenhet i oppgaver og tilhørende ulike løsninger, vil utfordringen for vurdererne være å komme til enighet om hva som skal vurderes i de ulike (5) fasene av opplæringen (Gessler & Herrera, 2015, s. 156). Et sentralt element ved vurderingssystemene er å se vurderingen i et livslangt læringsperspektiv, hvor det er et ønske om å fremme kompetanser som antas å understøtte framtidig læring og de generiske kompetanser. De tre vurderingssystemene har alle dette som formål, og MECVET-kriterier kan vel sies å tydeliggjøre det på en god måte. Håndtering av endringer og «lære og lære» vil være sentrale elementer i utviklingen av en yrkeskompetanse, noe som underbygges av Boud og Hawk (2003, s. 5) med deres syn om «sustainable assessment». Som et eksempel viste nemndsmedlem F2 gjennom veiledning og kommunikasjon med lærlingen betydningen av å lære underveis i fagprøven.

Denne diskusjonen av vurderingspraksiser til vurdererne viser det komplekse samspillet som påvirker og former vurderernes vurderinger. Funn om kompetansemål, kriterier, grenseoppganger, diskusjoner og samarbeid samt karakterer vil være med på å belyse forskningsspørsmålene.

#### 6.4 Vurderernes profesjonelle erfaringsbakgrunn

Funn viste at vurdererne lente seg til sin egen faglige bakgrunn i vurderingene. Med en kompleks yrkeskompetanse vil det være av betydning at vurdererne anvender sine egne erfaringer og sitt eget kjennskap til elektrikerfaget inn i vurderingene, noe Wolf (1995, s. 67) viser til. Det kan også relateres til O'Grady&Alwis (2012, s. 200) sin «professional body». Men det er forutsatt at man har en oppdatert forståelse av faget, og følger (3) yrkets utvikling (Gessler & Herrera, 2015, s. 156). Vurderernes erfaringsbakgrunn kommer til syne gjennom et engasjement i vurderingsarbeidet, noe Crisp (2012, s. 17) sin forskning støtter. Gjennom å anvende seg selv som fagperson viser vurdererne at de tar vurderingsjobben svært seriøst. De er opptatt av å gjøre et arbeid som ivaretar lærlingene i størst mulig grad gjennom anvendelse av sin yrkeskompetanse. Beckett (2008, s. 29) viser til at listen med kompetanser som skal vurderes må være forankret i praktiske vurderinger. En vurderer må være en del av en «community» hvor det hele tiden foregår endringer og justeringer som bidrar inn i vurderingsarbeidet. I HELVYRD-studien var det en forskjell mellom vurderingsfelleskapene.

Vurdererne i fagprøven og eksamen hadde lang fartstid og kjente mer til hverandre enn det raterene gjorde i MECVET. Det kan påvirke vurderingsprosessen. Nærheten til lærlingene vil også ha en betydning i den sammenheng, hvor førsteinntrykk vil kunne ha en påvirkende effekt i vurderingene.

Et annet funn var betydningen av om vurderernes erfaringsbakgrunn kom fra næringsliv eller skolen. Det fremkom kritiske røster blant prøvenemndsmedlemmene som hevdet at «skolefolk» ikke kunne bedømme en fagprøve. Det innebar om lærere hadde nok oppdatert kompetanse og om vektleggingen på faglig innhold er annerledes enn for de som kom fra næringslivet. Det framkom ulikt syn i vurderingene når vurdererne anvender egne erfaringer og deres forståelse av yrket. Enkelte besvarelser ble vurdert til at fremvist kompetanse ikke var nødvendig for å utføre jobben som elektriker, som for eksempel i hvilken grad beregninger skulle trekkes inn. I de tilfellene var uenigheten mellom vurdererne knyttet til bakgrunn, skole og næringsliv. Den (5) temporale delen (Gessler & Herrera, 2015, s. 156) av opplæringen (nevnt i forrige kapittel) kan også være et moment i den enkelte vurderers profesjonelle bakgrunn. Vurdererne må ta hensyn til hvor i opplæringsløpet lærlingene er, noe sensor E3 bemerket. Vil lærernes erfaringsbakgrunn bli mer verdsatt ved de skriftlige formene, som for eksempel ved eksamen og MECVET? Man kan stille et kritisk spørsmål til om vurdererne fra næringslivet kan ta med sine faglige erfaringer fra en arbeidskontekst og anvende dem direkte inn i en vurderingskontekst. Funn viste at vurdererne var åpne for faglige innspill fra lærlingene. De sensorene ved eksamen som var lærere uttrykte at vurderingsarbeidet var et moment for å holde seg faglig ajour. De mente at lærlingene kunne besitte kompetanse utover det sensorene hadde selv. At vurdererne ikke fremstår som «eksperter» men i større grad bygger på tillit til lærlingen, kan være med på å støtte det yrkesdidaktisk perspektiv som omhandler den tette (3) forbindelsen til elektrikerfaget (Gessler & Herrera, 2015, s. 156). Diskusjonen rundt vurderere fra skole eller næringsliv kan også bringe frem behovet for å se en miks av vurderere. Det kan gi en erfaringsutveksling som kan være betydningsfullt i vurderingene. Med ulike erfaringsbakgrunner kan det fremme økt grad av helhetlig tenkning rundt vurderingene og frembringe en bevisstgjøring, noe sensor E3 var innom. For stor grad av monokulturer kan virke hemmende. Som tidligere nevnt i kapittel 6.1 og 6.3 er vurderernes deltakelse i fellesskap med erfaringsdeling og diskusjoner er vesentlig for å utvikle en vurderingskultur som i størst mulig grad ivaretar gode vurderinger for å sikre kvaliteten til elektrikerfaget. Samtidig er det viktig å være bevisst at vurdererne kan fremstå



aggregerende, som Wolf (1995, s. 71) peker på. I den sammenheng er åpenhet viktig for å ivareta den faglige utviklingen.

Avslutningsvis vil vurderernes evne til å benytte yrkeslivet som en del av en referanseramme i vurderingen innebærer at de sammenligner og gjenkjenner. Dermed er det viktig å ivareta og verdsette vurderernes erfaringsbakgrunn. Det å gi rom for vurderernes skjønnsutøvelse er vesentlig for å ivareta vurderernes egen yrkeskompetanse inn i vurderingene. Med fokus på sterkt skjønn (Dworkin, 1978, s. 32) vil det fremme muligheten for å ta hensyn til en kompleks yrkeskompetanse, hvor det i større grad er rom for å legge til grunn for helhetlige vurderinger. Både Wolf (1995, s. 67) sin redegjørelse om de profesjonelle vurderernes holistiske modell og forskningen til Nore og Lahn (2015, s. 164) bygger opp under betydningen av skjønn som en del av grunnlaget for vurderinger. Men det er vesentlig at skjønn bygger på at den enkelte vurderer, som tidligere nevnt, er faglig oppdatert i elektrikerfaget. Flere eksempler fra vurdererne underbygde betydningen av det, slik blant annet sensor E3 uttrykte at det å ha sluttkompetansen med praktiske løsninger med seg i vurderingen var viktig. Det kan knyttes til Gessler og Herrera (2015, s. 153 og 156) sine perspektiver om både (1) læringsutbytte og (3) forming og utvikling av yrker (3).

Videre skal jeg i neste kapittel 6.5 syntetisere studiens funn om helhetlig vurdering i et yrkesdidaktisk perspektiv.

## 6.5 Studiens funn

Basert på diskusjonene i kapittel 5 og den sammenfattende diskusjonen om helhetlig vurdering i et yrkesdidaktisk perspektiv i kapittel 6.1-6.4, skal en syntetisering av studiens funn presenteres. Denne syntesen av funn og diskusjoner gjøres mer overordnet slik at studiens hovedkonklusjoner tydeliggjøres. Først presenteres en syntetisering av studiens funn (6.5.1), deretter noen betraktninger om kunnskapssyn og helhetlig vurdering (6.5.2) hvor det så avrundes med hvordan forskningsspørsmålene er besvart (6.5.3).

### 6.5.1 Syntetisering av studiens funn

Helhetlig vurdering av kompleks yrkeskompetanse er en øvelse som krever godt oppdaterte og erfarne vurderere. Vurderernes helhetlige vurderinger bygger i stor grad på deres egen rolle og erfaring som yrkesutøvere i elektrikerfaget (professional body). Yrkeslivet og yrkesutøvelsen trekkes aktivt inn som del av en vurderingsramme når vurdererne bruker yrkeskompetansen som elektriker. Det gir rom for å se helhetlig på besvarelsene til

lærlingene, men krever at yrkeskompetansen til vurdererne er oppdatert. På den annen side vil det være behov for en åpenhet i vurderingssystemene slik at vurdererne kan tilpasse oppgavene og vurderingskriteriene og derigjennom ta hensyn til ulike sider av en yrkeskompetanse. Bruk av skjønn i et samspill med formelle krav og retningslinjer for vurderingsarbeidet blir en del av vurderernes handlinger. Samtidig skjer dette i et vurderings- og tolkningsfellesskap med andre vurderere hvor samarbeid og diskusjon er en del av bevisstgjøringen om hva som er ønsket fagkompetanse og som de ønsker at lærlingene skal vise. I fellesskapet tolker vurdererne oppgavetekster, kompetansemål og vurderingskriterier, og diskuterer hva de skal legge vekt på i vurderingene. Som en del av vurderingsprosessen ligger også forventninger om faglige krav og diskusjoner vurderere imellom om hva som bør være det faglige «tyngdepunktet» i vurdering av lærlingenes besvarelser. I disse diskusjonene inngår også grenseoppganger mellom ulike roller og nivåer i fagområdet og hvordan det virker inn på tolkning av faglig innhold, kompetansemål og kriterier.

Samspillet mellom vurderernes egen fagkompetanse, kompetansemål og vurderingskriterier gir både muligheter og utfordringer for helhetlige vurderinger. Der mål og kriterier var for styrende og stramme, utfordret det vurdererne i å få gjort helhetlige vurderinger av lærlingenes besvarelser. Der vurdererne hadde et eierforhold til vurderingssystemene gjennom at de aktivt bidro i prosessene med oppgaveutforminger og kriterier, ga det større rom for å kunne gjøre helhetlige vurderinger. I et yrkesdidaktisk perspektiv må vurderingsformer, oppgavetekster, kompetansemål og kriterier kunne gjenspeile og gjenkjennes i elektrikeryrket. Det var vurdererne opptatt av. Bevisstheten om hva som kreves i elektrikeryrket i dag og hva som ser ut til å være utviklingen av yrket er påvirkende faktorer i vurderingsprosessen. Fagets innholdselementer er ikke en statisk liste, men diskuteres fortløpende både som et overordnet søkelys på elektrikeryrket og i tilknytning til de enkelte vurderingskontekstene. Uansett ligger innholdskomponenten el.-sikkerhet til grunn for all vurdering. De øvrige innholdselementene avhenger av konteksten og dermed innebærer det en pragmatisk tilnærming til helhetlig vurdering av en yrkeskompetanse påvirket både av vurderingssystemets representativitet og vurderingskonteksten.

Vurdererne har ulike strategier for å identifisere lærlingens yrkeskompetanse og gjøre helhetlige vurderinger av løsningsforslagene deres. Når utviklingen går i retning av åpne, casebaserte oppgaver må det stå i samsvar med hvordan kriterier utformes og anvendes i møte med en større spredning i lærlingenes løsninger og besvarelser. Helhetlig vurdering innebærer et samspill mellom vurderer og lærlingenes løsninger, hvor vurderernes fagkompetanse,

skjønn, eiforhold til vurderingssystemer og ulike strategier for vurdering er faktorer som spiller inn i vurderingsprosessen. Det kan gjenspeiles i mitt forslag til definisjon presentert i kapittel 3.1.9.

Helhetlig vurdering er å trekke inn ulike aspekter av en handling knyttet til en yrkesutøvelse i en konkret kontekst, der vurderer bruker sin erfaring, kompetanse og skjønn for i størst mulig grad å vurdere yrkeskompetansen opp mot de krav som stilles for utøvelsen av et yrke.

HELVYRD-studien har vist at vurderingssystemene påvirker vurderernes muligheter for helhetlig vurdering i et yrkesdidaktisk perspektiv.

For prøvenemndsmedlemmer som vurderer fagprøven gis det muligheter for å vurdere mer helhetlig gjennom lærlingenes muntlige og skriftlige presentasjoner samt observasjon av praktisk yrkesutøvelse. I tillegg er det mulig å bruke kompetansemål og kriterier (selv om de er forhåndsbeskrevet) på en fleksibel måte, ikke minst gjennom samarbeid og diskusjoner i et vurderingsfellesskap. På den annen side ser det ut til at oppgaveformer ved prøvestasjoner og bruken av prøvestasjon i seg selv begrenser mulighetene for helhetlige vurderinger.

Sensorene ved eksamen begrenses til lærlingenes skriftlige presentasjoner der lærlingenes skriveferdigheter også kunne påvirke vurderingene. Samtidig ble mulighetene for helhetlige vurderinger styrket gjennom utforming av casebaserte, åpne oppgaver og et tett samarbeid mellom sensorene i fortløpende utvikling og bruk av kriterier.

For raterne ved MECVET-testen fremstår de fyldige casebaserte oppgavetekstene og den stramme strukturen i vurderingsprosessen som minst egnet til å gjøre helhetlige vurderinger av en yrkeskompetanse.

Sett i et yrkesdidaktisk perspektiv, vil alle vurderingssystemene ha et potensiale til å være «suited to the task», men det kreves utvikling og tilpasninger. Samtidig vil det ikke være mulig å utvikle det perfekte vurderingssystem for kompleksiteten i en yrkeskompetanse. Som flere forskere peker på, er det ønskelig med flere og komplementerende vurderingssystemer.

### [6.5.2 Kunnskapssyn og helhetlig vurdering av yrkeskompetanse](#)

Før en oppsummerende konkludering av studiens forskningsspørsmål vil jeg trekke frem hvorvidt det er mulig å uttale seg, basert på studiens funn, hvilket syn vurdererne har på vurdering av yrkeskompetanse. Hvis vi bruker Bagnall og Hodge (2017) epistemologiske

retninger når vi skal forstå vurderernes kunnskapssyn, finner vi at disse kan ha kjennetegn som passer inn i alle de fire retningene i varierende grad og påvirket av mulighetene som ligger i det å gjøre vurderinger. Vurderere kan ha et konstruktivistisk syn på vurdering hvor de, uavhengig av vurderingssystem, etterstreber å se lærlingenes besvarelser i en helhet hvor det er et sterkt fokus på helheten hos lærlingen og om vedkommede kunne gjort en jobb som elektriker. Dette skjer i et «samspill» med de formelle retningslinjer. Det fremkommer at åpne casebaserte oppgaver kan underbygge for dette. Samtidig er det en bevissthet om begrensningene i hvert vurderingssystem hvor det komplementære forholdet mellom vurderingssystemene bidrar til å få et mer helhetlig blikk på yrkekompetansen. Det er et vesentlig funn selv om påvirkningen fra de tre vurderingssystemene i ulik grad påvirker vurderere i deres prosess. Sistnevnte kan trekkes inn i forhold til et mer sosiokulturelt syn. Vurdererne er opptatt av konteksten lærlingene skal operere i og at de er opptatt av at de kan vise kreativitet og kritisk tenkning i et yrke som er i sterk utvikling. Spesielt kan det sies å fremstå i de skriftlige vurderingene under eksamen og MECVET-testen. Men selv om fagprøven var mer låst til konkrete oppgaver hvor rammen påvirket en del så var det rom der også for å vise til utvikling. Det disiplinorienterte kom i mindre grad frem. Selvfølgelig lå det et faglig fokus, ikke i form av faginndelinger, men som en del av en helhet og hvordan det faglige anvendes inn i ulike kontekster og oppgaver. Innholdselementent el.-sikkerhet hadde en sentral plass, noe som kan føre oss mer over til det instrumentelle synet. Det instrumentelle ligger først og fremst i det å forholde seg til kompetansemål og kriterier. Alle vurderere mente kompetansemålene og kriteriene ikke måtte bli for instrumentelle. Det var enighet om at kompetansemål og kriterier skulle være med, men anvendelsen måtte gi rom for å anvende deres egen kompetanse og utøvelse som vurderer. Så vurderere var i mindre grad opptatt av å være instrumentelle i deres vurderinger, men igjen så fremstår MECVET-testen i større grad som instrumentell. Kort oppsummert var det ikke intensjonen i HELVYRD-studien å gå dypt inn på vurderernes «epistemologiske hverdagsoppfatninger», men å bruke noen av begrepene innen kunnskapsteori for å si noe om både vurderingssystemene og hvordan de ble brukt. Samlet sett var den rådende oppfatningen til vurdererne i HELVYRD studien at de i sine vurderinger av en kompleks yrkeskompetanse mest nærliggende hadde et konstruktivistisk syn, noe som kan sies å understøtte for helhetlige vurderinger. Avslutningsvis vil jeg kort nevne i kjølvannet av dette at Bagnall og Hodge (2017) har tatt opp mulighetene for å utvikle en ny epistemologi, som de omtaler som *situational*. Det har utgangspunkt i at man går vekk fra tanken om mer massive og større tilnærminger i et globalt syn, til å bli mer fokusert på skreddersøm og lokale tilnærminger og kontekster, men fortsatt i rammen av det globale.

Dette trenger mer teoretisering og utvikling og i den prosessen er tanken at den instrumentelle retningen kan ta til seg aspekter fra de andre retningene. Dette synet kan innebære en ytterligere kritikk av KOMET-modellen hvor stor-skalaundersøkelser og mer standardiseringer av vurderinger ligger til grunn.

### 6.5.3 Svar på forskningsspørsmålene

Hvordan har forskningsspørsmålene blitt belyst gjennom studiens funn?

Det første forskningsspørsmålet omhandlet det konseptuelle ved studien og lød:

*-Hvordan kan vi forstå helhetlig vurdering i et yrkesdidaktisk perspektiv?*

Helhetlig vurdering i et yrkesdidaktisk perspektiv handler om en variant av vurdering innenfor fag- og yrkesopplæring, og knyttes til behovet for å vurdere utførelsen av en kompleks helhetlig yrkesoppgave i ulike kontekster.

Yrkesdidaktikken omhandler først og fremst den tette tilhørigheten til yrkeslivet med yrkesutøvelsen og yrkesoppgaver, og som utgjør et utgangspunkt for opplæringen innen yrkesfag. Tilknyttet en vurderingssammenheng vil det være to faktorer. Det handler om utvikling av autentiske vurderingsoppgaver med utgangspunkt i elektrikerfaget, og det handler om å trekke yrkeslivet inn i vurderingsprosessene. De seks perspektivene som ble utledet om yrkesdidaktisk, med støtte i Gessler og Herrera (2015) seks forutsetninger, har gitt et ytterligere grunnlag for å forstå yrkesdidaktikkens innhold. Spesielt har perspektivene om læringsutbytte-forståelse, hvordan yrker er med på å forme den yrkesdidaktiske arenaen samt de globale påvirkningene og at det er ulike faglige nivåer bidratt til å belyse mine funn om helhetlig vurdering.

Helhetlig vurdering er utledet i bevissthet om det sammensatte kompetansebegrepet og det mer spesifikke begrepet yrkeskompetanse, hvor vurdering har sine utfordringer (Mulder, 2007). Valgt strategi om å lage et rammeverk med ulike aspekter ved helhetlig vurdering viser flere sider av en vurderers vurderinger og har resultert i en bred innsikt om hva begrepet kan bety. Helhetlig vurdering vil romme et syn på vurdering som handler om å fremme den faglige helheten. Definisjon av helhetlig vurdering kan sett i lys av funn i studien sies å både utfordre og støtte forståelsen av helhetlig vurdering, noe som vil bli belyst i de to neste forskningsspørsmålene. Betydningen av begrepet for vurdererne var at det handlet om å sikre at alle deler av en oppgave var med i en besvarelse.

Helhetlig vurdering i et yrkesdidaktisk perspektiv er en type vurdering som skjer i en bestemt kontekst med ulike påvirkende faktorer. Gjennom å belyse helhetlig vurdering i utvalgte yrkesdidaktiske perspektiver kommer det til syne muligheter, utfordringer og begrensninger. Spesielt den yrkesdidaktiske tilhørigheten og påvirkningen fra yrkeslivet vil være sentralt. Yrkenes stadige utvikling kommer tydelig frem, noe som preger både vurderingsfeltet og det yrkesdidaktiske feltet isolert sett, men også hvordan disse kan ses i sammenheng. Betyr det at helhetlig vurdering må beskrives i henhold til det enkelte faget? Helhetlig vurdering kan ha andre aspekter knyttet til det karakteristiske ved elektrikerfaget enn det vil i rørleggerfaget. Viktige momenter ved det enkelte fag kan bli utelatt hvor det særegne må ivaretas. Lovreguleringen av elektrikerfaget påvirker utførelsen og hvordan yrket kan bli vurdert. Dermed kan man snakke om en helhetlig fagvurdering på lik linje som man kan snakke om yrkesfagdidaktikken og tittelen på avhandlingen kan spisses til «*Helhetlig fagvurdering i et yrkesfagdidaktisk perspektiv*». Helhetlig vurdering i et yrkesdidaktisk perspektiv må ta utgangspunkt i de enkelte fag og blir dermed yrkesfagdidaktisk. Konkludert så må svaret på dette forskningsspørsmålet ses med det aktuelle faget i den aktuelle konteksten, i mitt tilfelle elektrikerfaget, og man skal være tydelig på at generalisering vil være utfordrende med tanke på det karakteristiske ved hvert fag.

Det andre forskningsspørsmålet omhandlet det deskriptive og innslaget av helhetlige vurderinger og lød:

*-I hvilken grad foretar vurdererne helhetlige vurderinger i de utvalgte vurderingssystemene?*

Graden av helhetlige vurderinger hos vurdererne var høy. Argumentet for det ligger i hvordan vurdererne tilnærmet seg elektrikerfagets innholdselementer og metodiske fremgangsmåte, bevisstheten om og bruken av deres egen faglige bakgrunn og yrkesutøvelsen og hvordan de forholdt seg til og anvendte vurderingskriterier.

Det som er med på fremme helhetlige vurderinger er vurderernes tilnærming til å vurdere besvarelser opp imot elektrikerfaget og yrkeslivet. Vurderernes evne til å imøtekomme alle besvarelser gjennom å finne det unike og særegne og å se det fra lærlingenes side og dermed ha lærlingens bakgrunn og potensiale som en del av den helhetlige vurderingen.

Samtidig er vurdererne klar over og kjenner til de formaliteter som gjelder ved vurdering, hvor de balanserer dette i vurderingsprosessen. Det at det gis rom for at vurdererne kan utvikle og anvende vurderingskriterier bidrar til helhetlig vurdering.

For å opprettholde en høy grad av helhetlig vurdering var det viktig med vurderingsfellesskaper. Både for å bli bevisstgjort og for å holde seg oppdatert i elektrikerfaget.

Deres sterke tilknytning til eget fagområde preger vurderernes vurderinger, og utgjør en sentral del ved helhetlig vurdering hvor deres skjønn og «professional body» blir en integrert del av den helhetlige vurderingen. Å inneha tillit til vurdererne i vurderingsprosessen innebærer også å tillate en sterk grad av skjønn.

Det tredje forskningsspørsmålet omhandlet det forklarende og påvirkningen fra vurderingssystemene og lød:

*-Hvilke faktorer påvirker vurderernes helhetlige vurdering av siste års lærlingbesvarelser?*

Flere faktorer påvirket vurderere i deres helhetlige vurderinger og først og fremst er det relatert til vurderingssystemenes form, sentrale retningslinjer og den pågående utviklingen i elektrikerfaget.

Vurderingssystemene påvirker vurdererne gjennom oppgavens utforminger, strukturer for vurderingsprosessene og de rammene som gis. Ulikheten mellom systemene er store, og dermed påvirkes vurdererne i ulik grad. Påvirkningen i MECVET-testen gjennom sin rigide struktur på vurderingsprosedyren gir i mindre grad et handlingsrom for både å tilpasse og å få frem det individuelle i lærlingens besvarelser. MECVET-kriteriene påvirket i for stor grad og begrenset den helhetlige vurdering. Kompetansemålene ved eksamen og fagprøven i sin nåværende form påvirker ikke i den grad og det gis et handlingsrom for tolkninger og tilpasninger og dermed i større grad et eieforhold og medvirkning i vurderingsprosessen.

Casebaserte oppgaver oppleves som tett innpå yrkesutøvelsen og i samsvar med elektrikerfaget og de gir rom for å imøtekomme lærlingenes egen besvarelse, noe den casebaserte eksamen var en bra representant for. For mye tekst preget MECVET-testen og sammen med den rigide strukturen for vurdering ble påvirkningen stor til ikke å kunne gjøre helhetlig vurderinger. En rekke forskere understøttet betydningen av åpenhet og ikke bli oppdelt i flere små deler, noe nettopp MECVET-testen gjør.

Elektrikerfagets utvikling er en faktor som påvirker vurdererne i stor grad. De var opptatt av å ha oppdatert kompetanse og innsikt noe som preget både synspunkter på oppgaveutforminger og anvendelse av vurderingskriterier hvor vurderingsfellesskapene var sentrale.

## 7. Avslutning og forslag til videre forskning

Jeg skal avslutte med noen refleksjoner om studiens bidrag, metodiske valg, teoretisk rammeverk og til slutt peke på noen mulige fremtidige studier på feltet.

### 7.1 Studiens bidrag

Tidligere forskning på vurdering i fag- og yrkesopplæringsfeltet er begrenset og stort sett tilknyttet et vurderingssystem og gjerne flere fag. HELVYRD-studien har bidratt med empirisk forskning på et felt innen vurdering som ikke er belyst i norsk fag- og yrkesopplæring hvor det er valgt å gå i dybden på et fag og flere vurderingsystemer. Studiens funn bidrar til å belyse en rekke forhold omkring både vurderernes handlinger og vurderingssystemer, og gir et detaljert bilde på det komplekse samspillet i vurderingen av en kompleks yrkeskompetanse. Studien har bidratt med empiri ved flere aspekter som uttrykker noe om mulighetene og utfordringene for vurdererne, og spesielt betydningen av deres tilknytning til yrkeslivet. Studiens funn om vurderingshandlingene har blitt belyst gjennom flere yrkesdidaktiske perspektiver, og har dermed bidratt til vise kompleksiteten og påvirkningen det har på vurdererne og deres helhetlige vurderinger.

Studios bidrag er også knyttet til å tre ulike vurderingssystemer hvor forskningen på vurdering har et komparativt perspektiv. Det perspektivet kan få frem begrensninger og muligheter i det enkelte vurderingssystem og kan bidra til å få øye på flere faktorer enn om man bare ser på et system alene. Spesielt, viser empirien i denne studien at ingen av systemene alene klarer å fange opp alle sider av en kompleks yrkeskompetanse. Studien viser derimot at det kan være en fordel å benytte komplementerende vurderingssystemer som kan brukes i ulike kontekster og ved ulike tidspunkt i et opplæringsløp.

Tidligere forskning har vist betydningen av vurderernes rolle i ulike vurderingssystemer. Denne studien bidrar med empiri som underbygger betydningen av vurderernes sentrale rolle og at den er selve kjernen i det å legge til rette for helhetlige vurderinger. Det å ha tillit til vurdererne samt et fleksibelt og tilpasset rammeverk som bygger oppunder deres faglige og vurderingstekniske kompetanse, må til skal en helhetlig vurdering av en helhetlig yrkeskompetanse vurderes.



## 7.2 Refleksjoner om det metodiske ved studien

Studien har en bred innfallsvinkel (eksplorerende), i og med at det ble gjort intervjuer av vurderere tilknyttet flere vurderingssystemer. I kapittel 4.8 ble det gjort noen oppsummerende betraktninger om kvaliteten på design og metode med støtte i Tracy (2010) sine «Eight «Big-Tent» Criteria for Excellent Qualitative Research», samt Kvale og Brinkmann (2015) sine syv stadier for validitet. Studiens validitet ses på som akseptabel. En ytterligere økning av validiteten kunne vært gjort ved å intervju flere vurderere, skape enda bedre rom for å gjennomføre intervjuene umiddelbart rett etter vurderingshandlingen og man kan utvikle intervjuguiden med spørsmål som kan gå dypere inn i vurderernes handlinger. Det handler om å skape tillit og respekt i intervjusituasjonen. Men for å være en første studie på denne delen av vurderingsfeltet så er det fremskaffede materialet rikt på detaljer som gir et godt bilde av situasjonen.

Min plassering i forskningsfeltet må ses i lys av min bakgrunn. Jeg har erfaring med deltakelse i begge de to norske vurderingssystemene og jeg har en elektrobakgrunn som gjør at jeg har en viss innsikt i elektrikerfaget. Det gir meg en forforståelse som det er nødvendig å være seg bevisst. Fordelen er at jeg allerede har god innsikt gjennom egne erfaringer om vurderingssystemenes funksjoner. Det gir meg et godt utgangspunkt for å stille spørsmål som er relevante for studien. Samtidig er det viktig å være klar over at min implisitte forståelse av vurderingens prosesser kan gjøre meg «blind» på vesentlige områder, noe som kan innebære at sentrale deler ikke blir nok belyst. Min rolle som forsker krever et åpent blikk, noe som kan være utfordrende hvis min forforståelse i for stor grad dominerer. Vesentlige elementer som jeg ser på som selvfølgeligheter blir ikke bli løftet frem. Men først og fremst mener jeg det er en fordel at jeg har hatt en god innsikt i forkant av det å gjennomføre en studie.

## 7.3 Refleksjoner om det teoretiske rammeverket

Mitt teoretiske rammeverk er delt i to hovedområder, yrkesdidaktikk og helhetlig vurdering (inkludert tidligere forskning). Yrkesdidaktikken er en mer overordnet ramme hvor helhetlig vurdering er en del av vurderingsfeltet som innorder seg under de didaktiske rammene. I ettertid kan jeg se at en ytterligere spissing av teorien mot helhetlig vurdering er mulig. Men mitt valg i denne studien var å gå bredt ut, siden feltet ikke er forsket mye på. Det var viktig å fange opp ulike aspekter som kunne belyse mine forskningsspørsmål. Å ha to hovedområder med teori krever en struktur som gir rom for at anvendelsen av dem i presentasjoner og diskusjoner kommer mest mulig til sin rett. Og det skjedde hovedsakelig i to faser. Den første

fasen var i kapittel 5, hvor teori om helhetlig vurdering ble anvendt i diskusjonen med ønske om å ha en nærhet til empiri. I den sammenfattende diskusjonen (kapittel 6), med støtte i Gessler og Herrera (2015) sine forutsetninger, ble funn om helhetlig vurdering ytterligere belyst gjennom det yrkesdidaktiske perspektivet. Her trakk jeg også inn noen nye perspektiver på vurdering. Todelingen av det teoretiske rammeverket krever at funn blir gjennomgående samtidig som det blir en kondensering og syntetisering av funn. Det var en utfordrende øvelse. Mitt valg om å gå bredt ut ga muligheter for å se studiens forskningsspørsmål i et bredt spekter, men jeg ser i ettertid at man kunne kommet dypere inn i diskusjonen med færre vinklinger. Det vil gi spennende muligheter for videre forskning og som blir tatt opp i neste avsnitt.

#### 7.4 Forslag til tema for nye forskningsstudier

HELVYRD-studiens forskningsspørsmål har gitt studien en eksplorerende innfallsvinkel. Det har bidratt til en innledende innsikt om vurderernes helhetlige vurderinger innenfor et bestemt yrke. For videre forskning vil det i første omgang være naturlig å utvide til flere fagområder. Det vil ytterligere bidra til en større innsikt i hva helhetlig vurdering innebærer, og om det har mer lik eller ulik betydning i de enkelte yrkene. Det gir et bredere grunnlag for å ha et komparativt perspektiv. I denne sammenheng vil en ytterligere forståelse av yrkesdidaktikken bli en del av det.

Det vil også være hensiktsmessig å benytte flere typer av vurderingssystemer. De utvalgte i HELVYRD-studien har blitt omtalt som summative vurderingsordninger. I den teoretiske diskusjon ble det vist til sammenhengen mellom formativ og summativ vurdering. Det vil være interessant å gjøre mer forskning på hvordan vurdering i de to fasene i større grad kan ha en sammenheng, og hvordan det påvirker den helhetlige vurderingen. Det omhandler både forutsigbarhet og rettferdighet overfor den lærende som skal vurderes. Samtidig vil det være viktig å forske mer på hvordan ulike vurderingsformer påvirker læringsprosessen og hvordan disse understøtter både elevers og lærlingers læring.

Et område med behov for ytterligere forskning er utforming og anvendelse av kriterier for vurdering i ulike kontekster. Det må sees i sammenheng med vurderernes skjønn og sentrale retningslinjer og det samspillet som oppstår. I hvilken grad skal kriterier ha betydning og i hvilken grad skal de styre vurderingene? Går det an å lage tydelige bestemte kriterier innen fag- og yrkesopplæring, eller må de tilpasses den enkelte konteksten hvor vurderer har et pragmatisk forhold til mer åpne og fleksible kriterier? HELVYRD-studien har vist gjennom

både teori og empiri at mer forskning og diskusjon om denne delen av vurdering er nødvendig. Samspillet mellom egen faglig bakgrunn, kompetansemål og kriterier, samt skjønn, skaper en sammensatt kontekst for hvordan vurdering gjennomføres. Mer forskning kan gjøres igjennom metoder som think-aloud- og stimulated recall-prosesser. Det er vesentlig med tanke på at vurderingen setter premisser for hvordan opplæring organiseres. Det hjelper ikke med aldri så mange sentralt styrte systemer og metoder, da det til syvende og sist er vurdererne det gjelder og som besitter den kompetanse som skal anvendes i vurderingen. Det er det vesentlig å få mer innsikt i.

Skriftlighet blant lærlingene framstod som viktig. Et element som ikke ble tatt med i analysen er betydningen av å skrive ved hjelp av PC. Både ved eksamen (enkelte lærlinger med spesiell tillatelse), fagprøven og MECVET-testen ble PC anvendt. Spesielt ved eksamen ble det uttrykt det ikke alltid var like greit. Lærlingene med PC kunne levere de flotteste tegninger og dokumentasjoner, men lærlingene som benyttet penn og papir kunne streve noe mer. De som benyttet PC kunne få minst én karakter bedre enn de uten. Det påvirker vurderingen fordi oppsettet blir mer ryddig, og dermed ser man lettere helheten. Nemndsmedlemmene uttalte også at bruk av PC kan være en fordel fordi besvarelsen kan fremstå som mer ryddig. Ved MECVET-testen måtte alle lærlingene bruke PC ved sine besvarelser, men det ble i mindre grad diskutert fordeler og ulemper. Dette er en del av skrivingens betydning som det kan forskes mer på

Et siste felt med behov for mer forskning er faglig innhold, noe som er og skal være det sentrale i vurderingen. Går det an å definere konkrete faglige områder som kommer tydelig frem, eller vil det være behov for større grad av pragmatisme? I den sammenheng vil fremtidens yrker ha betydning og hvordan disse utformes. Hvordan skal vurdererne forholde seg til en stadig utvikling innenfor elektrikerfaget? Kompetansebegrepet gir en form for retning når det gjelder hva kommende arbeidstakere skal besitte, og det utvikles stadig mer i en kompleks retning. Det får også betydning for hvordan vurdering gjennomføres og påvirkes. Hva innebærer en helhetlig vurdering med stadig mer kompleksitet i yrkesutøvelsene? Går det an å ha helhetlige vurderinger som vil romme alt det en fremtidig yrkesutøver skal ha av kompetanse? Eller vil Wolf (1995) sin påstand fortsatt gjelde, om at vi alltid bare vil kunne vurdere bruddstykker av en yrkeskompetanse? Det må det forskes mer på.

Feltet er utforsket, spesielt i en norsk kontekst. Det vil være grunnlag for å gjøre mange flere studier om vurdering som kan belyse ulike sider ved en sentral og viktig del av norsk fag- og

yrkesopplæring, spesielt når fagfornyelsen kommer med nye læreplaner i 2020. Hvordan det vil ha betydning for vurdering generelt og helhetlig vurdering spesielt, vil være interessant å forske på. Innledningsvis ble det tatt opp at nye metoder for vurdering ble utviklet, hvor blant annet storskalaundersøkelser var en av dem. Achtenhagen og Winther (2014, s. 292) sin studie om utvikling av fremtidige storskala vurderingsformer i fag- og yrkesopplæring kan være et viktig bidrag. Noen av funnene der, spesielt i forbindelse med å utvikle valide yrkesoppgaver, vil også kunne anvendes inn i andre former for vurdering. G. Haaland (2018) har i sin bok sett på flere aspekter av vurdering i fag- og yrkesopplæringen i forhold til helhetlig yrkeskompetanse. Her ligger det mange spennende vinklinger for videre utforskning om vurderingens rolle.

Avslutningsvis kan man stille spørsmål om hvilke helheter som kan inngå i helhetlig vurdering. Innledningsvis i kapittel 1.2 ble det stilt noen spørsmål om hva helhetlig kan innebære. I etterkant av studien kan man stille følgende spørsmål. Er det forholdet mellom skole og arbeidslivet? Er det faget som helhet? Den enkelte lærling som helhet? Vurderingen i seg selv som en helhet? Vurderingsformene som en helhet? Disse områdene kan gi ytterligere bidrag til videre forskning på helhetlig vurdering.

## Litteraturliste

- Aarkrog, V. (2004). En model for samspillet mellem skole og praktik i vekselluddannelserne. I P. Bottrup & C. Helms Jørgensen (Red.), *Læring i et spændingsfelt: Mellem uddannelse og arbejde* (s. 169-180). Roskilde: Roskilde universitetsforlag.
- Aarkrog, V. (2005). Learning in the workplace and the significance of school-based education: A study of learning in a Danish vocational education and training programme. *International Journal of Lifelong Education*, 24(2), 137-147. <https://doi.org/10.1080/02601370500056268>
- Aarkrog, V. (2010). *Fra teori til praksis: Undervisning med fokus på transfer*. København: Munksgaard.
- Achtenhagen, F. & Winther, E. (2014). Workplace-based competence measurement: Developing innovative assessment systems for tomorrow's VET programmes. *Journal of Vocational Education & Training*, 66(3), 281-295. <https://doi.org/10.1080/13636820.2014.916740>
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223. <https://doi.org/10.3200/TCHS.78.5.218-223>
- 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>
- Baethge, M. & Arends, L. (2009). *Feasibility Study VET-LSA: A comparative analysis of occupational profiles and VET programmes in 8 European countries - International report* (2. utg.). Bonn: Bundesministerium für Bildung und Forschung.
- Bagnall, R. G. & Hodge, S. (2017). Using an epistemological perspective to understand competence-based vocational and professional education. I M. Mulder (Red.), *Competence-based vocational and professional education, technical and vocational education and training: Issues, concerns and prospects* (s. 125-144). Switzerland: Springer International Publishing.
- Baumeler, C. (2019). Competence-based vocational education and training and its cultural context sensitivity. *European Education*, 51(1), 1-15. <https://doi.org/10.1080/10564934.2017.1344503>
- Baumert, J. (2009). International comparisons: Strengths and weakness and how to overcome the weaknesses. I F. K. Oser, U. Renold, E. G. John, E. Winther & S. Weber (Red.), *VET boost: Towards a theory of professional competencies: Essays in honor of Frank Achtenhagen* (s. 25-37). Rotterdam: Sense.
- Beckett, D. (2008). Holistic competence: Putting judgements first. *Asia Pacific Education Review*, 9(1), 21-30. <https://doi.org/10.1007/BF03025822>

- Beckett, D. & Hager, P. (2002). *Life, work and learning: Practice in postmodernity*. New York: Routledge.
- Biemans, H., Wesselink, R., Gulikers, J., Schaafsma, S., Verstegen, J. & Mulder, M. (2009). Towards competence-based VET: Dealing with the pitfalls. *Journal of Vocational Education & Training*, 61(3), 267-286. <https://doi.org/10.1080/13636820903194682>
- Bjørndal, B. & Lieberg, S. (1978). *Nye veier i didaktikken?: En innføring i didaktiske emner og begreper*. Oslo: Aschehoug.
- Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education*, 26(3-4), 407-419. <https://doi.org/10.1080/713688540>
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167. <https://doi.org/10.1080/713695728>
- Boud, D. & Hawke, G. (2003). Changing pedagogy: Vocational learning and assessment. *OVAL Research Working paper*, (03-17), 1-8.
- Brevik, B. (2014). *LEGO & læring: En kvalitativ studie av elektrofaglæreres bruk av LEGO Mindstorms som læringsverktøy i norsk videregående skole* (Doktorgradsavhandling). Universitetet i Oslo, Oslo.
- Calderhead, J. (1981). A psychological approach to research on teachers' classroom decision-making. *British Educational Research Journal*, 7(1), 51-57. <https://doi.org/10.1080/0141192810070105>
- Carlgrén, I., Forsberg, E. & Lindberg, V. (2009). *Perspektiv på den svenska skolans kunskapsdiskussion* (Rapport (Centrum för studier av skolans kunskapsinnehåll i praktiken) nr. 7/2009). Stockholm: Stockholms universitets förlag.
- Charmaz, K. (2014). *Constructing grounded theory* (2. utg.). Los Angeles: Sage.
- Choy, S., Wärvik, G.-B. & Lindberg, V. (2018). Considerations for the integration of students' experiences. I S. Choy, G.-B. Wärvik & V. Lindberg (Red.), *Integration of vocational education and training experiences* (s. 345-365). Singapore: Springer Nature Singapore.
- Colley, H. & Jarvis, J. (2007). Formality and informality in the summative assessment of motor vehicle apprentices: A case study. *Assessment in Education: Principles, Policy & Practice*, 14(3), 295-314. <https://doi.org/10.1080/09695940701591883>
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4. utg.). Los Angeles: Sage.
- Crisp, V. (2012). An investigation of rater cognition in the assessment of projects. *Educational Measurement: Issues and Practice*, 31(3), 10-20. <https://doi.org/10.1111/j.1745-3992.2012.00239.x>
- Dale, E. L. & Wærness, J. I. (2007). *Vurdering og læring i en elevaktiv skole*. Oslo: Universitetsforlaget.

- Deichman-Sørensen, T., Høst, H., Michelsen, S., Nore, H., Olsen, O. J. & Hagen Tønder, A. (2011). *Prøvenemndenes arbeid med fag- og svenneprøver: En undersøkelse av fem fag* (FAFO-rapport nr. 2011:11). Oslo: Fafo.
- De nasjonale forskningsetiske komiteene. (2016, 31. mai). Generelle forskningsetiske retningslinjer. Hentet fra <https://www.etikkom.no/forskningsetiske-retningslinjer/Generelle-forskningsetiske-retningslinjer/>
- Denzin, N. K. & Lincoln, Y. S. (2018). *The SAGE handbook of qualitative research* (5. utg.). Los Angeles: Sage.
- Digital21. (2018). [Strategi]. Hentet fra [http://digital21.no/wp-content/uploads/2018/09/Digital21\\_strategi\\_2018.pdf](http://digital21.no/wp-content/uploads/2018/09/Digital21_strategi_2018.pdf)
- Dobson, S. (2010). Nasjonale og internasjonale utfordringer innen elev- og lærlingvurdering. I S. Dobsen & R. Engh (Red.), *Vurdering for læring i fag* (s. 24-36). Kristiansand: Høyskoleforlaget.
- Dworkin, R. (1978). *Taking rights seriously*. London: Duckworth.
- Ecclestone, K. (2010). Assessment in vocational education. I P. Peterson, E. Baker & B. McGaw (Red.), *International encyclopedia of education* (3. utg., bd. 3, s. 337-341). Amsterdam: Elsevier.
- Eisner, E. W. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York: Macmillan Publ. Co.
- Ellström, P.-E. (1992). *Kompetens, utbildning och lärande i arbetslivet: Problem, begrepp och teoretiska perspektiv*. Stockholm: Publica.
- Engelsen, B. U. (1998). *Fagprøve og fagforståelse med fragmenter i et bilde av fagopplæring i arbeidslivet* (Rapport nr. 6 1998). Oslo: Universitetet i Oslo.
- Engh, K. R., Dobson, S. & Høihilder, E. K. (2007). *Vurdering for læring*. Kristiansand: Høyskoleforlaget.
- Eraut, M. (1998). Concepts of competence. *Journal of Interprofessional Care*, 12(2), 127-139. <https://doi.org/10.3109/13561829809014100>
- Eraut, M. (2004). A wider perspective on assessment. *Medical Education*, 38(8), 803-804. <https://doi.org/10.1111/j.1365-2929.2004.01930.x>
- Flick, U. (2007). *Managing quality in qualitative research*. London: Sage.
- Forskrift til opplæringslova. (2006). (FOR-2006-06-23-724). Hentet fra <https://lovdata.no/dokument/SF/forskrift/2006-06-23-724>
- Friche, N. (2010). *Erhvervsskolens evalueringspraksis: Intentioner bag, anvendelse og virkning af evaluering i erhvervsuddannelserne* (Doktorgradsavhandling). Aalborg University, Aalborg.

- Friche, N. (2013). Evaluering som logik og praksis i erhvervsuddannelser, (14), 14-19.  
<https://doi.org/10.17896/UCN.cepra.n14.103>
- Gessler, M. & Herrera, L. M. (2015). Vocational didactics: Core assumptions and approaches from Denmark, Germany, Norway, Spain and Sweden. *International Journal for Research in Vocational Education and Training*, 2(3), 152-160.  
<https://doi.org/10.13152/IJRVET.2.3.1>
- Gessler, M. & Howe, F. (2015). From the reality of work to grounded work-based learning in German vocational education and training: Background, concept and tools. *International Journal for Research in Vocational Education and Training*, 2(3), 214-238. <https://doi.org/10.13152/IJRVET.2.3.6>
- Gillis, S. & Griffin, P. (2014). Principles underpinning graded assessment in VET: A critique of prevailing perceptions. *International Journal of Training Research*, 3(1), 53-78.  
<https://doi.org/10.5172/ijtr.3.1.53>
- Gipps, C. (1999). Socio-cultural aspects on assessment. *Review of Research in Education*, 24, 355-392. Hentet fra <http://www.jstor.org/stable/1167274>
- Goodlad, J. (1979). *Curriculum inquiry: The study of curriculum practice*. New York: McGraw-Hill Book Company.
- Greatorex, J. & Sütő, I. W. M. (2008). What do GCSE examiners think of 'thinking aloud'? Findings from an exploratory study. *Educational Research*, 50(4), 319-331.  
<https://doi.org/10.1080/00131880802499720>
- Grimen, H. & Molander, A. (2008). Profesjon og skjønn. I A. Molander & L. I. Terum (Red.), *Profesjonsstudier* (s. 179-196). Oslo: Universitetsforlaget.
- Gulikers, J. M., Bastiaens, T. J. & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research & Development*, 52(3), 67-86
- Gundem, B. B. (2011). *Europeisk didaktikk: Tenkning og viten*. Oslo: Universitetsforlaget.
- Haaland, G. (2018). *Vurdering i yrkesfag: Helhetlig yrkeskompetanse*. Oslo: Pedlex.
- Haaland, G. & Nilsen, S. E. (2013). *Læring gjennom praksis: Innhold og arbeidsmåter i yrkesopplæringen: En grunnbok i yrkesdidaktikk*. Oslo: Pedlex
- Haaland, G. S., Nore, H. & Vagle, I. (2009). Vurdering for og av læring i yrkesfag. I S. Dobsen, A. B. Eggen & K. Smith (Red.), *Vurdering, prinsipper og praksis: Nye perspektiver på elev- læringsvurdering* (s. 214-234). Oslo: Gyldendal akademisk.
- Haasler, B. & Erpenbeck, J. (2008). Assessing vocational competences. I F. Rauner & R. Maclean (Red.), *Handbook of technical and vocational education and training research* (s. 766-774). Netherlands: Springer Netherlands.
- Hager, P. (2017). The integrated view on competence. I M. Mulder (Red.), *Competence-based vocational and professional education* (s. 203-228). [https://doi.org/10.1007/978-3-319-41713-4\\_10](https://doi.org/10.1007/978-3-319-41713-4_10)



- Hager, P. & Gonczi, A. (1996). What is competence? *Medical Teacher*, 18(1), 15-18.  
<https://doi.org/10.3109/01421599609040255>
- Hager, P., Gonczi, A. & Athanasou, J. (1994). General issues about assessment of competence. *Assessment & Evaluation in Higher Education*, 19(1), 3-16.  
<https://doi.org/10.1080/0260293940190101>
- Halvorsen, K. (2008). *Å forske på samfunnet* (5. utg.). Oslo: Cappelen Akademiske Forlag.
- Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*, 33(1), 15-28. <https://doi.org/10.1016/j.stueduc.2007.01.003>
- Hayton, G. & Wagner, Z. (1997). Performance assessment in vocational education and training. *Australian and New Zealand Journal of Vocational Education Research*, 6(1), 69-85.
- Helle, L. (2007). *Læringsrettet vurdering*. Oslo: Universitetsforlaget.
- Hellne-Halvorsen, E. B. (2014). *Skrivepraksiser i yrkesfaglige utdanningsprogrammer* (Doktorgradsavhandling). Universitetet i Oslo, Oslo.
- Hellne-Halvorsen, E. B., Lahn, L. C. & Nore, H. (In press). Writing competences in norwegian vocation education and training - how students and apprentices express their professional competences. *Written Communication*.
- Herrera, L. M. (2015). Transitions and diversity in didactics: An exploration searching for implications for vocational education and training, 2(3), 161-169.  
<https://doi.org/10.13152/IJRVET.2.3.2>
- Hiim, H. (2012). Relevant, praksisbasert yrkesutdanning. I B. Aamotsbakken (Red.), *Ledelse og profesjonsutøvelse: Uløselig sammenvevde begreper?* (s. 123-137). Oslo: Universitetsforlaget
- Hiim, H. & Hippe, E. (2001). *Å utdanne profesjonelle yrkesutøvere: Yrkesdidaktikk og yrkeskunnskap*. Oslo: Gyldendal akademisk.
- Hiim, H. & Hippe, E. (2009). *Undervisningsplanlegging for yrkesfaglærere* (3. utg.). Oslo: Gyldendal akademisk.
- Hoy-Mack, P. (2005). Workplace assessment in New Zealand: stated intentions and realisations. *International Journal of Training Research*, 3(1), 79-95.  
<https://doi.org/10.5172/ijtr.3.1.79>
- Imsen, G. (2009). *Lærerens verden: Innføring i generell didaktikk* (4. utg.). Oslo: Universitetsforlaget.
- Inglar, T. (2009). *Erfaringslæring og yrkesfaglærere: En kvalitativ studie* (Doktorgradsavhandling). Høgskolen i Akershus, Oslo.
- Jacobs, P. A. (2016). *The potential of the comet competence diagnostic model for the assessment and development of occupational competence and commitment in technical*

- vocational education and training* (Doktoravhandling). Bremen University, Bremen. Hentet fra <http://nbn-resolving.de/urn:nbn:de:gbv:46-00105542-11>
- Johannessen, A., Christoffersen, L. & Tufte, P. A. (2016). *Introduksjon til samfunnsvitenskapelig metode* (5. utg.). Oslo: Abstrakt forlag.
- Johansson, S. (2008). Några relevanta begrepp inom yrkeslärande. I S. Johansson (Red.), *Yrkesläroarbildning inom SÄL III-prosjektet* (s. 139-156). Vällingby, Sverige: Elander.
- Johnson, M. (2008a). Assessing at the borderline: Judging a vocationally related portfolio holistically. *Issues in educational research*, 18(1), 26-43.
- Johnson, M. (2008b). Grading in competence-based qualifications – is it desirable and how might it affect validity? *Journal of Further and Higher Education*, 32(2), 175-184. <https://doi.org/10.1080/03098770801979183>
- Jones, A. (1999). The place of judgement in competency-based assessment. *Journal of Vocational Education & Training*, 51(1), 145-160. <https://doi.org/10.1080/13636829900200073>
- Kirke-, utdannings- og forskningsdepartementet. (1996, 28. mars). Vidaregåande opplæring etter Reform 94. Hentet fra <https://www.regjeringen.no/no/dokumentarkiv/regjeringen-brundtland-iii/kuf/veiledninger/1996/Vidaregaande-opplaring-etter-Reform-94/id87404/>
- Kuczera, M., Brunello, G., Field, S. & Hoffman, N. (2008). *Learning for Jobs: OECD reviews of vocational education and training: Norway*. Hentet fra [https://www.researchgate.net/publication/242464778\\_Learning\\_for\\_Jobs\\_OECD\\_Reviews\\_of\\_Vocational\\_Education\\_and\\_Training\\_Norway](https://www.researchgate.net/publication/242464778_Learning_for_Jobs_OECD_Reviews_of_Vocational_Education_and_Training_Norway)
- Kunnskapsdepartementet. (2006). ... og ingen sto igjen: Tidlig innsats for livslang læring (St.meld. nr. 16 (2006-2007)). Oslo: Departementet.
- Kunnskapsdepartementet. (2011). *Motivasjon - Mestring - Muligheter: Ungdomstrinnet* (Meld. St. 22 (2010-2011)). Oslo: Departementet.
- Kunnskapsdepartementet. (2013). *På rett vei: Kvalitet og mangfold i fellesskolen* (Meld. St. 20 (2012-2013)). Oslo: Departmentet.
- Kunnskapsdepartementet. (2016). *Fag - Fordypning - Forståelse: En fornyelse av kunnskapsløftet* (Meld. St. 28 (2015-2016)). Oslo: Departmentet.
- Kunnskapsdepartementet. (2017a, 25. august). Digitaliseringstrategi for grunnopplæringen 2017-2021: Plan/strategi. Hentet fra <https://www.regjeringen.no/no/dokumenter/framtid-fornyelse-og-digitalisering/id2568347/>
- Kunnskapsdepartementet. (2017b). *Overordnet del - verdier og prinsipper*. Hentet fra <https://www.regjeringen.no/contentassets/37f2f7e1850046a0a3f676fd45851384/overordnet-del---verdier-og-prinsipper-forgrunnopplaringen.pdf>

- Kunnskapsdepartementet. (2019, 23. april). Kunnskapsløftet. Hentet fra <https://www.regjeringen.no/no/tema/utdanning/grunnopplaring/kunnskapsloftet/id534689/>
- Kuusela, H. & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis.(Statistical Data Included). *American Journal of Psychology*, 113(3), 387-404. <https://doi.org/10.2307/1423365>
- Kvale, S. & Brinkmann, S. (2015). *Det kvalitative forskningsintervju* (3. utg.). Oslo: Gyldendal akademiske.
- Kvernbekk, T. (2005). *Pedagogisk teoridannelse: Insidere, teoriformer og praksis*. Bergen: Fagbokforlaget.
- Lahn, L. C. & Nore, H. (2019). Large-scale studies of holistic professional competence in Vocational Education and Training (VET): The case of Norway. *International Journal for Research in Vocational Education and Training*, 6(2), 132-152. <https://doi.org/10.13152/IJRVET.6.2.2>
- Lai, E. R. (2011). Performance-based assessment: Some new thoughts on an old idea. *Pearson Education*, (20), 1-4.
- Laming, D. (2004). *Human judgment: The eye of the beholder*. London: Thomson Learning.
- Larsson, S. (2009). A pluralist view of generalization in qualitative studies. *International Journal of Research & Method in Education*, 32(1), 25-38. <https://doi.org/10.1080/17437270902759931>
- Lier, A. R. (2017). *Fra arbeidsoppgave til læringsoppgave: En studie av yrkesfaglærernes arbeid med faget Prosjekt til fordypning PTF innenfor programområdet Teknologi og industriell produksjon (TIP)* (Doktoravhandling). Roskilde Universitet, Roskilde.
- Lindberg, V. (2003). Learning practices in vocational education. *Scandinavian Journal of Educational Research*, 47(2), 157-179. <https://doi.org/10.1080/00313830308611>
- Lounsbery, J. L. & Pittenger, A. L. (2011). Instructional design and assessment: Case-based Assessment in an Online Nursing Pharmacotherapy Course. *American Journal of Pharmaceutical Education*, 75(6), 1-8. <https://doi.org/10.5688/ajpe756114>
- Lum, G. (2013). Competence: A tale of two constructs. *Educational Philosophy and Theory*, 45(12), 1193-1204. <https://doi.org/10.1080/00131857.2013.763593>
- Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British Educational Research Journal*, 29(6), 861-878. <https://doi.org/10.1080/0141192032000137349>
- Lynggaard, K. (2015). Dokumentanalyse. I S. Brinkmann & L. Tanggaard (Red.), *Kvalitative metoder: En grundbog* (2. utg., s. 153-167). København: Hans Reitzel.
- Lyngsnes, K. M. & Rismark, M. (2014). *Didaktisk arbeid* (3. utg.). Oslo: Gyldendal akademisk.

- Markussen, E. (2007). Reform 94 lever videre - men videregående svikter hver femte ungdom. I H. Hølleland (Red.), *På vei mot Kunnskapsløftet: Begrunnelser, løsninger og utfordringer* (s. 84-111). Oslo: Cappelen akademisk forlag.
- Mulder, M., Weigel, T. & Collins, K. (2007). The concept of competence in the development of vocational education and training in selected EU member states: A critical analysis. *Journal of Vocational Education & Training*, 59(1), 67-88. <https://doi.org/10.1080/13636820601145630>
- Nilsson, L. (1992). Fagdidaktikk i yrkespedagogisk perspektiv. I L. Mjelde & A.-L. H. Tarrou (Red.), *Arbeidsdeling i en brytningstid: Yrkespedagogiske utfordringer i skole og arbeidsliv* (s. 96-110). Oslo: Ad notam Gyldendal.
- Nilsson, L. (2000). Den glömde arbeidsoppgiften. I Utbildningsdepartementet, *Samverkan mellan skola och arbetsliv: Om möjligheterna med lärande i arbete* (s. 226-264, Ds 2000:62). Stockholm: Utbildningsdepartementet.
- Nore, H. (2015). Re-contextualizing vocational didactics in Norwegian vocational education and training. *International Journal for Research in Vocational Education and Training*, 2(3), 182-194. <https://doi.org/10.13152/IJRVET.2.3.4>
- Nore, H. & Lahn, L. C. (2015). Resultat kvalitet uttrykt gjennom fag- og svenneprøver. I H. Høst (Red.), *Kvalitet i fag- og yrkesopplæring: Sluttrapport* (s. 145-171, NIFU-rapport nr. 14/2015). Oslo: Nordisk institutt for studier av innovasjon, forskning og utdanning.
- NOU 2008:18. (2008). *Fagopplæring for fremtida*. Oslo: Kunnskapsdepartementet.
- NOU 2014:7. (2014). *Elevenes læring i fremtidens skole: Et kunnskapsgrunnlag*. Oslo: Kunnskapsdepartementet.
- NOU 2015:8. (2015). *Fremtidens skole: Fornyelse av fag og kompetanser*. Oslo: Kunnskapsdepartementet.
- NOU 2018:2. (2018). *Fremtidige kompetansebehov I*. Oslo: Kunnskapsdepartementet.
- NOU 2019:2. (2019). *Fremtidige kompetansebehov II*. Oslo: Kunnskapsdepartementet.
- Nyen, T. & Tønder, A. H. (2014). *Yrkesfagene under press*. Oslo: Universitetslaget.
- Nærings- og fiskeridepartementet. (2017). *Industrien - grønnere, smartere og mer nyskapende* (Meld. St. 27 (2016-2017)). Oslo: Departementet.
- O'Grady, G. & Alwis, W. A. M. (2012). Holistic assessment and problem-based learning. I G. O'Grady, E. H. J. Yew, K. P. L. Goh & H. G. Schmidt (Red.), *One day, one problem: An approach to problem-based learning* (s. 187-212). Singapore: Springer Singapore.
- Olsen, O. J. (2009). *Feasibility Study VET-LSA: Rapport til KD om norsk deltakelse*. Hentet fra <http://docplayer.me/7341349-Feasibility-study-vet-lsa-rapport-til-kd-om-norsk-deltakelse.html>
- Opplæringslova. (1998). Lov om grunnskolen og den videregående opplæringa (LOV-1998-07-17-61). Hentet fra <https://lovdata.no/lov/1998-07-17-61>

- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research and Evaluation*, 13(4), 1-11.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4. utg.). Los Angeles: Sage.
- Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy* (Rev. utg.). Chicago: The University of Chicago Routledge & Kegan Paul.
- Rafoss, T. W. & Witsø, H. (2014). Fagenes krav og lovens bokstav – En kvantitativ undersøkelse av prøvenemndene på Agder. *Norsk pedagogisk tidsskrift*, 98(2), 78-92.
- Rauner, F. (2014). *Überprüfen beruflicher Handlungskompetenz zum Zusammenhang von Prüfen und Kompetenzdiagnostik*. Bremen: Bremen University, FG Berufsbildungsforschung (IBB).
- Rauner, F., Haasler, B., Heinemann, L. & Grollmann, P. (2009). *Messen beruflicher kompetenzen: Bd. 1. Grundlagen und Konzeption des KOMET-Projektes* (2. utg.). Berlin: LIT Verlag.
- Rauner, F., Heinemann, L., Maurer, A., Haasler, B., Erdwien, B. & Martens, T. (2013). *Competence development and assessment in TVET (COMET): Theoretical framework and empirical results*. Dordrecht: Springer Science.
- Rokkones, K. & Fiskvik, K. (2010). Vurdering av elektrofag og restaurant- og matfag. I S. Dobsen & R. Engh (Red.), *Vurdering for læring i fag* (s. 189-200). Kristiansand: Høyskoleforlaget.
- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. I G. Joughin (Red.), *Assessment, learning and judgment in higher education* (s. 1-15). [https://doi.org/10.1007/978-1-4020-8905-3\\_4](https://doi.org/10.1007/978-1-4020-8905-3_4)
- Sannerud, A. R. (2005). *Læring på byggeplassen - utopi eller realitet? Et aksjonsforskningsprosjekt i byggbransjen* (Doktgradsavhandling). Roskilde Universitetscenter. Forskerskolen Livslang Læring, Roskilde.
- Sannerud, A. R. (2006). Yrkeskompetanse - en begrepsanalyse. I E. Askerøi & O. Eikeland (Red.), *Som gjort, så sagt? Yrkeskunnskap og yrkeskompetanse* (bd. 13, s. 219-241). Lillestrøm: Høgskolen i Akershus.
- Schommer-Aikins, M., Unruh, S. & Morpew, J. (2015). Epistemological belief congruency in mathematics between vocational technology students and their instructors. *Journal of Education and Training Studies*, 3(4), 137-145. <https://doi.org/10.11114/jets.v3i4.859>
- Segers, M., Martens, R. & van den Bossche, P. (2008). Understanding how a case-based assessment instrument influences student teachers learning approaches. *Teaching and teaching education*, 24(7), 1751-1764. <https://doi.org/10.1016/j.tate.2008.02.022>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14. <https://doi.org/10.3102/0013189X029007004>

- Silverman, D. (2014). *Interpreting qualitative data* (5. utg.). Los Angeles: Sage.
- Strand, T. & Kvernbekk, T. (2000). Problems of educational models and their use. *Nordisk Pedagogik*, 20(1), 1-12.
- Strauss, A. L. & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2. utg.). Thousand Oaks, Calif.: Sage.
- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W. & Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation*, 32(3), 202-222.  
<https://doi.org/10.1016/j.stueduc.2006.08.002>
- Sütő, I. (2012). A critical review of some qualitative research methods used to explore rater cognition. *Educational Measurement: Issues and Practice*, 31(3), 21-30.  
<https://doi.org/10.1111/j.1745-3992.2012.00240.x>
- Tanggaard, L. & Elmholdt, C. (2008). Assessment in practice: An inspiration from apprenticeship. *Scandinavian Journal of Educational Research*, 52(1), 97-116.  
<https://doi.org/10.1080/00313830701786719>
- Tanggaard, L. (2004). Evaluering som regulering af læring. *Arbejdsliv*, 6(4), 57-75.
- Tarrou, A.-L. H. (2004). Yrkesdidaktikk - kunnskapsstatus og forskningsbehov. I N. forskningsråd (Red.), *Kunnskapsstatus for forskningsprogrammet KUPP: Kunnskapsutvikling i profesjonsutdanning og profesjonsutøving* (s. 54-69). Oslo: Norges forskningsråd.
- Tarrou, A.-L. H. (2005). Danning i yrkesfagene i skolen og i lærerutdanningen. I K. Børhaug, A.-B. Fenner & L. Aase (Red.), *Fagenes begrunnelser: Skolens fag og arbeidsmåter i danningperspektiv* (s. 213-224). Bergen: Fagbokforlaget.
- Thagaard, T. (2013). *Systematikk og innlevelse: En innføring i kvalitativ metode* (4. utg.). Bergen: Fagbokforlaget.
- Thronsen, I., Hopfenbeck, T. N., Lie, S. & Dale, E. L. (2009). *Bedre vurdering for læring*. Oslo: Universitetet i Oslo.
- Tigelaar, D. E. H. & van der Vleuten, C. P. M. (2014). Assessment of professional competence. I S. Billett, C. Harteis & H. Gruber (Red.), *International handbook of research in professional and practice-based learning* (s. 1237-1270).  
[https://doi.org/10.1007/978-94-017-8902-8\\_45](https://doi.org/10.1007/978-94-017-8902-8_45)
- Tracy, S. J. (2010). Qualitative Quality: Eight “Big-Tent” Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, 16(10), 837-851.  
<https://doi.org/10.1177/1077800410383121>
- Tsagalidis H. (2008). *Därför fick jag bara Godkänt: Bedömning i karaktärsämnen på HR-programmet* (Doktorsavhandling). Stockholms universitet, Stockholm.

- Utdannings- og forskningsdepartementet. (2004). *Kultur for læring* (St. meld. nr. 30 (2003-2004)). Oslo: Departementet.
- Utdanningsdirektoratet. (2009). *Sluttrapport: Oppdragsbrev nr. 6 - 2007 om tiltak knyttet til individvurdering i skole og fag- og yrkesopplæring*. Hentet fra [https://www.udir.no/globalassets/upload/forskning/5/bedre\\_vurderingspraksis\\_sluttrapport\\_til\\_kd.pdf](https://www.udir.no/globalassets/upload/forskning/5/bedre_vurderingspraksis_sluttrapport_til_kd.pdf)
- Utdanningsdirektoratet. (2013). *Læreplan i elektrikerfaget Vg3 / opplæring i bedrift*. Hentet fra <http://data.udir.no/kl06/ELE3-02.pdf>
- Utdanningsdirektoratet. (2016a). Læreplanverket for Kunnskapsløftet. Hentet fra <https://www.udir.no/Udir/PrintPageAsPdfService.ashx?pid=112431&epslanguage=no>
- Utdanningsdirektoratet. (2016b, 18. mai). Å forstå kompetanse. Hentet fra <https://www.udir.no/laring-og-trivsel/lareplanverket/forsta-kompetanse/>
- Utdanningsdirektoratet. (2017a). Elektrofag. Hentet fra <https://www.udir.no/Udir/PrintPageAsPdfService.ashx?pid=112516&epslanguage=no>
- Utdanningsdirektoratet. (2017b). *Retningslinjer for samarbeid – SRY, faglige råd og Udir*. Hentet fra <https://www.udir.no/Udir/PrintPageAsPdfService.ashx?pdfld=128453>
- Utdanningsdirektoratet. (2018). Yrkesfag - rettleiing til sentralt gitt skriftleg eksamen. Hentet fra <https://www.udir.no/Udir/PrintPageAsPdfService.ashx?pid=111437&epslanguage=no>
- Utdanningsdirektoratet, Faglig råd for elektro. (2016). *Utviklingsredegjørelse 2016: Utdanningsprogram for elektro*. Hentet fra <https://fagligerad.files.wordpress.com/2015/12/frel.pdf>
- van der Vleuten, C., Sluijsmans, D. & Joosten-ten Brinke, D. (2017). Competence assessment as learner support in education. I M. Mulder (Red.), *Competance-based vocational and professional education* (s. 607-630). [https://doi.org/10.1007/978-3-319-41713-4\\_28](https://doi.org/10.1007/978-3-319-41713-4_28)
- Vaughan, C. (1991). Holistic assessment: What goes on in the raters' mind? I L. Hamp-Lyons (Red.), *Assessing second language writing in academic contexts* (s. 111-125). Norwood: Ablex.
- Vestøl, J. M. (2008). Didaktiske modeller i lærerutdanningen. *Acta Didactica Norge*, 2(1), 1-20.
- Williams, M. & Bateman, A. (2003). *Graded assessment in vocational education and training: An analysis of national practice, drivers and areas for policy development*. Leabrook, Australia: National centre for vocational education research.
- Wolf, A. (1995). *Competence-based assessment*. Buckingham: Open University Press.
- Wolf, A. (2001). Copmpetence-based assessment. I J. Raven & J. Stephenson (Red.), *Competence in the Learning Society* (s. 453-466). New York: Peter Lang Publishing.

Wood, T. J. (2014). Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education*, 19(3), 409-427.  
<https://doi.org/10.1007/s10459-013-9453-9>



## Vedlegg

### Vedlegg 1 Intervjuguider for vurderere i de tre utvalgte vurderingssystemene

#### **Prøvenemndsmedlemmer Fagprøver**

##### **Intervju av prøvenemndsmedlemmer etter vurderte besvarelser**

Intervju en og en rett etter at alle prøvene er avlagt. Deretter to og to.

Går igjennom kandidat for kandidat hvor informant uttrykker sine umiddelbare tanker om sin vurdering. Dokumentasjon (oppgavetekst pluss annen informasjon rundt prøven og kandidatens besvarelse) for hele fagprøven ligger foran oss

##### **Innledende**

*La oss se nærmere på denne besvarelsen. Her har vi prøven til x. Den fikk bestått/bestått meget godt/ikke bestått.*

*Hvis du ser tilbake til hva du tenkte da du observerte og snakket med kandidaten, hva la du vekt på i din vurdering?*

*Hva så du etter?*

*Hva er viktig?*

*Noe spesielt du la merke til?*

*Planleggingsdelen?*

*Gjennomføringsdelen og den praktiske løsningen?*

*Egenvurderingsdelen?*

*Dokumentasjonsdelen?*

*Hvorfor denne karakteren? Enighet i nemnda? Var du i tvil?*

*Var det noe som manglet? Var det noe som utmerket seg? Hvordan oppdaget du det?*

*Hva skulle evt til for å gått opp en karakter? Eller ned en karakter?*

Ved intervju to og to gikk vi igjen gjennom de ulike kandidatene. Spurte om hva de tenkte.

Lede inn på samtalen med å referere til hva en av kandidatene har sagt, og fortsette videre der ifra. Guiden over er fortsatt rammen for denne samtalen også.

## Oppfølgingsintervju en og en

Det gjennomføres noen dager i etterkant av første runde. Den første delen av oppfølgingen handler om presiseringer og utdypinger i forhold til det første intervjuet med skreddersydde spørsmål. Det ble utviklet stikkord og enkle setninger som jeg benyttet i intervjuet med den enkelte.

Den andre delen er mer generelle spørsmål med fokus på hvordan nemndsmedlemmene går frem i deres vurderingsarbeid og rammene rundt fagprøven.

- Individuelle «spørsmål»/punkter:

- F1 Fire lærlinger: En med bestått meget godt, to med bestått og en med ikke bestått.
- F2 Fire lærlinger: En med bestått meget godt, to med bestått og en med ikke bestått.
- F3 Tre lærlinger: To bestått og en ikke bestått.
- F4 Tre lærlinger: To bestått og en ikke bestått
- F5 Fire lærlinger: Tre bestått meget godt og en bestått.
- F6 Fire lærlinger: Tre bestått meget godt og en bestått. (F6 og F1 samme person)

- Spørsmål knyttet til gjennomføring og rammer:

*Hvordan går du frem når dere vurderer kandidatene?*

*Så du på vurderingskriteriene underveis? Hvordan brukte du disse? Kompetansemålene?*

*Hva vurderes underveis? Knyttet til alle delene og er det noe vektlegging?*

*Notater? Skjema?*

*Hva skal til for en i IB, MB og B? Hva legger du i det?*

*Hva legger du i helhetlig vurdering?*

*Sammenligner du lærlinger med hverandre?*

*Hvordan opplever du oppgavene og oppgavetekstene?*

*Er det noe ved denne prøveformen du mener har noen sterke/svake sider?*

*Andre elementer du vil legge til?*

## **Ratere MECVET-tester**

### **Oppfølgingsintervju av ratere (en og en) etter innlevert vurdering og lydfiler:**

Dette gjennomføres noen dager i etterkant av innlevert materiale fra ratere. Har hørt igjennom lydfilene og gått igjennom vurderingsskjemaene.

Intervjuet en og en rater og gikk igjennom kandidat for kandidat. Raterene skal få uttrykke sine tanker om vurderingen. Den første delen med spørsmål vil være rettet inn imot de konkrete besvarelsene til kandidatene. Anvendte de innledende spørsmålene for alle samt noen individuelle spørsmål. Dokumentasjon for hele MECVET-testen ligger foran oss (oppgaveteksten, besvarelse, løsningsrom, vurderingsskjemaer (enkeltvis og omforent)). Avslutningsvis kommer de mer generelle spørsmålene om gjennomføring og rammer.

### **Innledende med enkelte individuelle spørsmål**

*La oss se nærmere på denne besvarelsen. Her har vi prøven til x. Her ser vi vurderingsskjema.*

*Hvis du ser tilbake til hva du tenkte da du vurderte denne besvarelsen, hva la du vekt på i din vurdering?*

*Hva så du etter?*

*Hva er viktig?*

*Noe spesielt du la merke til? Noe som utmerket seg? Hvordan oppdaget du det?*

*Var det noe som manglet? Hva måtte mer til for å heve nivået?*

*Hvordan vil du begrunne nivået du har vurdert denne besvarelsen til?*

*Var du i tvil under vurderingen? Evt på hva? Ift til den andre rater?*

- Enkelte individuelle spørsmål

M3

M4

M1

M2

- Spørsmål knyttet til gjennomføring og rammer:

*Hvordan går du frem når du vurderer kandidaten/besvarelsen? (Din egen fagbakgrunn og identitet).*

*Gjør du deg noen notater?*

*Hvordan bruker du «vurderingskriteriene» underveis? Hvordan opplever du de (representative? savner?) og vurderingsskjema? Hvordan bruker du løsningsrommet? Hvordan opplever du det?*

*På hvilket grunnlag vurderte du nivået på besvarelsen under de ulike kriteriene?*

*Hvordan opplever du disse «kriteriene» kontra kompetansemål i læreplan?*

*Legger du i Ikke oppfylt- dårlig oppfylt – nesten oppfylt – oppfylt?*

*Hva legger du i helhetlig vurdering?*

*Hvordan opplever du oppgavene og oppgavetekstene? Relevans? Mulighet for å løse dem innenfor de rammene som er gitt? Nivå for elever/lærlinger/fagarbeidere?*

*Er det noe ved denne prøveformen du mener har noen sterke/svake sider?*

*Hvis du sammenligner din måte å vurderer besvarelsene på denne typen oppgaver med hvordan du vurderer prestasjonene på fagprøve- og på teoretisk prøve, hva er likheter og hva er forskjeller?*

*Andre elementer du vil legge til?*

## Sensorer skriftlig eksamen

### Intervju av sensorer (en og en) etter vurderte besvarelser:

Intervjut en og en sensor rett etter deres sensormøte. Går igjennom kandidat for kandidat (de utvalgte i hele skalaen om mulig).

Målet er at sensorene skal få uttrykke sine tanker om vurderingen de har gjort. Den første delen med spørsmål vil være rettet inn imot de konkrete besvarelsene til kandidatene.

Dokumentasjon for hele eksamen ligger foran oss (oppgaveteksten pluss annen informasjon rundt prøven og dokumentasjonen til kandidaten)

### Innledende

Ser på kandidatens konkrete besvarelse:

*La oss se nærmere på denne besvarelsen. Her har vi prøven til x. Vurderingen ble X.*

*Hvis du ser tilbake til hva du tenkte da du vurderte denne besvarelsen, hva la du vekt på i din vurdering?*

*Hva så du etter?*

*Hva er viktig?*

*Noe spesielt du la merke til? Noe som utmerket seg? Hvordan oppdaget du det?*

*Var det noe som manglet?*

*Var du i tvil under vurderingen? Evt på hva? (Enighet mellom sensorer)?*

*Hvordan vil du begrunne nivået og karakteren du har vurdert denne besvarelsen til?*

*Hva skulle evt til for å gått opp eller ned en karakter?*

*Andre ting du vil si nå i den første runden?*

### Oppfølgingsintervju (en og en)

Dette gjennomføres noen dager i etterkant av første runde hvor strategien svitsjer noe. Den første delen av dette handler om presiseringer og utdypinger i forhold til det første intervjuet med enkelte mer skreddersydde spørsmål. Jeg har gått igjennom dette (hørt igjennom lydfilene og sett på skriftlig materiale).

Den andre delen er noe mer generelle spørsmål som er litt utover det som ble fokusert på i første runde, med mer fokus på hvordan sensorene går frem i deres vurderingsarbeid og noe mer om rammene rundt den skriftlige eksamen.

- Individuelle spørsmål

E2 Tre lærlinger med karakter en 5, en 2 og en 3.

E4 Tre lærlinger med karakter en 5, en 2 og en 3

E3 Tre lærlinger med karakter en 2, en 6 og en 4

E1 Tre lærlinger med karakter en 2, en 6 og en 4

Spørsmål knyttet til gjennomføring og rammer:

*Hvordan går du frem når du vurderer kandidaten/besvarelsen?*

*Har dere vurderingskriterier? Bruker du de underveis? Hvordan evt brukte du disse?*

*Kompetansemålene i læreplan? Vektlegging?*

*Gjør du deg noen notater? Skjema?*

*Hva legger du i helhetlig vurdering?*

*Hvordan opplever du oppgavene og oppgavetekstene? Relevans? Mulighet for å løse dem innenfor de rammene som er gitt? Nivået?*

*Er det noe ved denne prøveformen du mener har noen sterke/svake sider?*

*Hva skal til for å få 1-6? Hva legger du i det? Sammenligner med andre?*

*Hvis du vurderer andre typer av oppgaver og besvarelser, og sammenligner det med din måte å vurdere denne besvarelsene på, hva er likheter og hva er forskjeller?*

*Andre elementer du vil legge til?*

## Kvittering fra NSD

**Norsk samfunnsvitenskapelig datatjeneste AS**  
NORWEGIAN SOCIAL SCIENCE DATA SERVICES



Harald Hårfages gate 29  
N-5007 Bergen  
Norway  
Tel: +47-55 58 21 17  
Fax: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org nr. 985 321 884

Roger Bakken  
Institutt for yrkesfaglærerutdanning Høgskolen i Oslo og Akershus  
Postboks 4 St. Olavs plass  
0130 OSLO

Vår dato: 24.04.2014

Vår ref: 38063 / 3 / 1B

Deres dato:

Deres ref:

### TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 10.03.2014. All nødvendig informasjon om prosjektet forelå i sin helhet 23.04.2014. Meldingen gjelder prosjektet:

<i>38063</i>	<i>HELVYRD - Helhetlig vurdering for profesjonslæring i et yrkesdidaktisk perspektiv</i>
<i>Behandlingsansvarlig</i>	<i>Høgskolen i Oslo og Akershus, ved institusjonens øverste leder</i>
<i>Daglig ansvarlig</i>	<i>Roger Bakken</i>

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i melde skjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 30.09.2016, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Katrine Utaaker Segadal

Inga Brautaset

Kontaktperson: Inga Brautaset tlf: 55 58 26 35

Vedlegg: Prosjektvurdering

*Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.*

*Afdelingskontorer / District Offices:*

OSLO: NSD, Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11. [nsd@iuh.no](mailto:nsd@iuh.no)  
TRONDHEIM: NSD, Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07. [kyre.swarva@svt.ntnu.no](mailto:kyre.swarva@svt.ntnu.no)  
TROMSØ: NSD, SVE, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 43 36. [radmar@svt.uib.no](mailto:radmar@svt.uib.no)



Meldingen gjelder et doktorgradsprosjekt, der formålet er å studere utvalgte vurderingssystemer innen fag- og yrkesopplæringen, for å se nærmere på hvordan man gjennomfører vurdering av yrkeskompetanse. Hvilke kriterier legger de vekt på for vurdering av yrkeskompetanse, og i hvilken grad er denne vurderingen helhetlig og har en yrkesvaliditet?

Prosjektet er en del av et større forskningsprosjekt støttet av Norsk Forskningsråd, MECVET - Measuring Competence in Vocational Education and Training.

#### INFORMASJON OG SAMTYKKE

Utvalget vil bestå av medlemmer i fagnemda og prøvenemda, samt sensorer. De informeres skriftlig og muntlig om prosjektet og samtykker til deltakelse.

Informasjonsskrivet er stort sett godt utformet, men vi legger til grunn at følgende endres:

- Det må tilføyes at forskergruppen i MECVET vil ha tilgang på innsamlede opplysninger.
- Dersom informantene vil kunne gjenkjennes indirekte i publikasjoner, bør dette stå eksplisitt i skrevet.
- Kontaktopplysninger til stipendiat må tilføyes.

Revidert informasjonsskriv skal sendes til [personvernombudet@nsd.uib.no](mailto:personvernombudet@nsd.uib.no) for utvalget kontaktes.

#### DATAMATERIALETS INNHOLD

Data innhentes ved:

- lydopptak av observasjon og intervju av nemndsmedlemmene i fagnemnda under oppgavekonstruksjon,
- lydopptak av observasjon og intervju av nemndsmedlemmene i fagnemnda/sensorer ved sensur av skriftlige eksamensoppgaver,
- lydopptak av observasjon og intervju av medlemmer i prøvenemnd ved oppgavekonstruksjon,
- lydopptak av observasjon og intervju av medlemmer i prøvenemnd/sensorer ved vurdering av fagprøve,
- Det er også aktuelt å gjøre observasjon under avlegging av fagprøven (offentlig arrangement), men det skal da ikke tas lydopptak.

Personvernombudet legger til grunn at taushetsplikten ikke er til hinder for at stipendiaten kan observere nemdenes/sensorenes vurdering av konkrete fagprøver/eksamensoppgaver. Vi anbefaler at stipendiaten avklarer dette skriftlig med rett instans (trolig utdanningsinstitusjonen).

Det tas høyde for at lydopptak fra vurdering av fagprøve/eksamensoppgaver kan inneholde identifiserende opplysninger om kandidaten (tredjeperson). Fokus fra observasjonene er på nemdas arbeid, ikke på kandidaten. Det registreres kun opplysninger som er nødvendig for prosjektets formål. Tredjepersonsopplysningene vil være av mindre omfang og ikke sensitive. Dersom kandidatens navn fremgår, bør disse slettes fra opptakene. Opplysningene om tredjeperson skal anonymiseres i publikasjon og ved prosjektslutt. Så fremt



personvernulempen for tredjeperson reduseres på denne måten, kan prosjektleder unntas fra informasjonsplikten overfor tredjeperson, fordi det anses uforholdsmessig vanskelig å informere.

Dokumentanalysen gjelder eksamensoppgaver (ikke besvarelser) og styringsdokumenter uten personopplysninger. Denne delen av prosjektet omfattes dermed ikke av meldeplikten.

#### DATASIKKERHET

Personvernombudet legger til grunn at forsker etterfølger Høgskolen i Oslo og Akershus sine interne rutiner for datasikkerhet. I datamaterialet skal navn erstattes med kode som viser til en koblingsnøkkel som lagres nedlåst og adskilt fra øvrige opplysninger. Dersom personidentifiserende opplysninger skal lagres på privat pc/mobile enheter, bør opplysningene krypteres tilstrekkelig.

Forskere i MECVET (Lahn, Deichman-Sørensen, Nore) vil ha tilgang til data.

Det kan bli aktuelt å bruke en ekstern transkriberingsassistent som databehandler. Høgskolen i Oslo og Akershus skal da inngå skriftlig avtale med transkriberingsassistenten om hvordan personopplysninger skal behandles, jf. personopplysningsloven § 15. For råd om hva databehandleravtalen bør inneholde, se Datatilsynets veileder: <http://www.datatilsynet.no/Sikkerhet-internkontroll/Databehandleravtale/>. Personvernombudet ber om kopi av avtalen for arkivering (sendes: [personvernombudet@msd.uib.no](mailto:personvernombudet@msd.uib.no)).

#### PROSJEKTSLUTT

Forventet prosjektslutt er 30.09.2016. Ifølge prosjektmeldingen skal innsamlede opplysninger da anonymiseres. Anonymisering innebærer å bearbeide datamaterialet slik at ingen enkeltpersoner kan gjenkjennes. Det gjøres ved å:

- slette direkte personopplysninger (som navn/koblingsnøkkel)
- slette/omskrive indirekte personopplysninger (identifiserende sammenstilling av bakgrunnsopplysninger som f.eks. bosted/arbeidssted, alder og kjønn)
- slette lydopptak

Vi gjør oppmerksom på at også databehandler (transkriberingsassistenten) må slette personopplysninger tilknyttet prosjektet i sine systemer.

# Endringsmeldinger

Norsk samfunnsvitenskapelig datatjeneste AS  
NORWEGIAN SOCIAL SCIENCE DATA SERVICES



Harald Hørfagres gate 29  
N-5007 Bergen  
Norway  
Tel: +47-55 58 21 17  
Fac: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org.nr. 985 321 884

Roger Bakken  
Institutt for yrkesfaglererutdanning  
Høgskolen i Oslo og Akershus  
Postboks 4 St. Olavs plass  
0130 OSLO

Vår dato: 27.03.2015

Vår ref: 38063/MBLR

Deres dato:

Deres ref:

## BEKREFTELSE PÅ ENDRING

Vi viser til prosjektet:

38063 *HELVYRD - Helbetylig vurdering for profesjonslæring i et yrkesdidaktisk perspektiv*

Personvernombudet har tidligere vurdert og bekreftet flere endringer i prosjektet, jf. våre eposter 13.10.14 og 31.10.14:

- Det foretas ikke observasjoner, kun intervju.
- Informantene vil, etter samtykke, kunne gjenkjennes indirekte i publikasjoner.
- Dato for forventet avslutning av doktorgradsprosjektet endret til 31.12.2017. Datamaterialet med personopplysninger skal da overføres til MECVET-prosjektet for videre forskning innen det angitte formålet, etter samtykke. Opplysningene anonymiseres når MECVET er avsluttet 31.03.2022.

Personvernombudet har nå mottatt melding om nye endringer i eposter 15.01.15, 19.01.15 og 06.02.15:

- Forskningen vil omhandle elektrikerfaget, ikke industrimekanikerfaget.
- Det skal likevel foretas enkelte observasjoner med feltnotater.
- Intervjuguiden for prøvenemdsmedlemmer som vurderer fagprøver er endret, slik at hovedfokus er på deres vurderinger av kandidatens fagprøver, ikke på oppgavekonstruksjon.
- Kandidaters oppgavebesvarelser/dokumentasjon vil inngå som del av dokumentasjonen og danne utgangspunkt for intervju med nemdsmedlemmene.

Personvernombudet har, etter vurdering av endringene og avklaring av enkelte spørsmål, følgende kommentarer til endringene.

- Fra observasjonene registreres kun anonyme opplysninger.
- Oppdatert informasjonsskriv (innsendt 26.03.15) vurderes godt utformet.
- Om kandidatene registreres opplysninger som, i følge forsker, vanskelig kan tilbakeføres til den enkelte. Utdanningsetaten har gitt tillatelse til at forsker får tilgang til besvarelser i anonymisert form.

Afdelingskontorer / District Offices:

OSLO: NSD, Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11, nsd@iuh.no  
TRONHEIM: NSD, Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07, kym.svarwa@it.ntnu.no  
TROMSØ: NSD, HSL, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 61 53, sokk.anderassen@uit.no

Personvernombudet legger til grunn at prosjektopplegget for øvrig er uendret. Vi vil ved planlagt avslutning for doktorgradsprosjektet, 31.12.2017, rette en henvendelse vedrørende status for behandling av personopplysninger.

Ta gjerne kontakt dersom noe er uklart.

Vennlig hilsen

  
Katrine Utaaker Segadal

  
Inga Bråutaset

# **Informasjonsskriv for prøvenemndsmedlemmer fagprøven**

## **Forespørsel om deltakelse i forskningsprosjektet**

### **HELVYRD**

#### **Bakgrunn og formål**

HELVYRD (Helhetlig vurdering for profesjonslæring i et yrkesdidaktisk perspektiv) prosjektet er et doktorgradsstudie som er en del av et større forskningsprosjekt støttet av Norsk Forskningsråd, MECVET (Measuring Competence in Vocational Education and Training). MECVET vil utvikle og prøve ut et modell-basert verktøy for måling av yrkeskompetanse. Verktøyet er basert på en modell utviklet i Bremen, Tyskland. I HELVYRD prosjektet skal jeg blant annet sammenligne vurderinger gjort etter den tyske modellen opp mot nemndsmedlemmer som vurderer fag- og svenneprøver. Faget som følges er elektriker.

Personene som deltar i undersøkelsen er valgt ut med bakgrunn i deres roller i de ulike nemndene de deltar i samt fagbakgrunn. I forhold til MECVET prosjektet følger jeg oppsatt design i forhold til utvalget av personer.

#### **Hva innebærer deltakelse i studien?**

Studien vil se nærmere på vurdering i nemndene. Dette innebærer at jeg vil gjennomføre dokumentanalyse av oppgaver til fagprøven og tilhørende dokumenter samt dokumentasjon fra lærlingens fagprøve, og intervjuer i prøvenemndene av medlemmene i forhold til vurderingen av besvarelsene. Det vil også bli gjennomført noen observasjoner underveis i gjennomføringen av fagprøven. Som støtte for mine intervjuer vil jeg benytte lydbånd-opptaker

#### **Hva skjer med informasjonen om deg?**

Dataene lagres på en forsvarlig måte og gjøres ikke tilgjengelig for andre enn meg og forskergruppen i MECVET. Alle personopplysninger vil bli behandlet konfidensielt. Dersom du godtar lydbånd-opptak og annen informasjon fra deg benyttes i forskningssammenheng, er det fint om du signerer under.

I utgangspunktet skal ikke personene som deltar i undersøkelsen kunne bli direkte gjenkjent i publikasjoner, da dette skal anonymiseres så langt det lar seg gjøre. Samtidig er det offentlig å vite hvem som sitter i de ulike nemndene i ulike fag i ulike fylker, og fagmiljøene kan være små og gjennomsluktige. Så for de involverte kan de bli gjenkjent indirekte.

HELVYRD prosjektet avsluttes den dagen undertegnede har disputert og avhandlingen publisert (i løpet av 2017). Personopplysninger og data vil overføres til MECVET-prosjektet for videre forskning innenfor det angitte formålet. Datamaterialet anonymiseres når MECVET er fullført, 31.03.2022.

### **Frivillig deltakelse**

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn.

Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert.

Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med undertegnede (se under).

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

### **Samtykke til deltakelse i studien**

Jeg har mottatt informasjon om studien, og er villig til å delta

-----  
(Signering, dato)

Jeg samtykker til å delta i intervju

Med vennlig hilsen

Roger Bakken  
Stipendiat  
Høgskolen i Oslo og Akershus  
Fakultet for Lærerutdanning og internasjonalisering  
Institutt for Yrkesfaglærerutdanningen  
Tlf: 95997770  
Mail: rbakken@hioa.no

Leder for MECVET-prosjektet:  
Leif Lahn  
Høgskolen i Oslo og Akershus  
Fakultet for Lærerutdanning og internasjonalisering  
Institutt for Yrkesfaglærerutdanningen  
Mail: leifchristian.lahn@hioa.no

# Informasjonsskriv for sensorer eksamen

## Forespørsel om deltakelse i forskningsprosjektet

### HELVYRD

#### Bakgrunn og formål

HELVYRD-prosjektet (Helhetlig vurdering for profesjonslæring i et yrkesdidaktisk perspektiv) er et doktorgradsstudie som er en del av det større forskningsprosjekt MECVET (Measuring Competence in Vocational Education and Training)s som er støttet av Norsk Forskningsråd. MECVET vil utvikle og prøve ut et modell-basert verktøy for måling av yrkeskompetanse. Verktøyet er basert på en modell utviklet i Bremen, Tyskland. I HELVYRD prosjektet skal jeg blant annet sammenligne vurderinger gjort etter den tyske modellen opp imot vurderinger som gjøres av sensorer på sentralgitt skriftlig eksamen i yrkesfag. Faget som følges er elektrikerfaget.

Personene som deltar i undersøkelsen er valgt ut med bakgrunn i deres roller som sensor for elektrikerfaget.

#### Hva innebærer deltakelse i studien?

Studien vil se nærmere på vurdering av eksamensbesvarelsene. Dette innebærer at jeg vil gjennomføre dokumentanalyse av eksamensoppgavene og tilhørende dokumenter. Dokumentasjon fra lærlingens besvarelse vil også være tilgjengelig som en støtte under intervjuene av sensorene i forhold til vurderingen av besvarelsene. Som støtte for mine intervjuer vil jeg benytte lydbånd-opptaker.

#### Hva skjer med informasjonen om deg?

Dataene lagres på en forsvarlig måte og gjøres ikke tilgjengelig for andre enn meg og forskergruppen i MECVET. Alle personopplysninger vil bli behandlet konfidensielt. Dersom du godtar lydbånd-opptak og annen informasjon fra deg benyttes i forskningssammenheng, er det fint om du signerer under.

I utgangspunktet skal ikke personene som deltar i undersøkelsen kunne bli direkte gjenkjent i publikasjoner, da dette skal anonymiseres så langt det lar seg gjøre. Samtidig er det offentlig å vite hvem som sitter som sensorer i de ulike fag i ulike fylker, og fagmiljøene kan være små og gjennomsiktige. Så for de involverte kan de bli gjenkjent indirekte.

HELVYRD prosjektet avsluttes den dagen undertegnede har disputert og avhandlingen publisert (i løpet av 2017). Personopplysninger og data vil overføres til MECVET-prosjektet for videre forskning innenfor det angitte formålet. Datamaterialet anonymiseres når MECVET er fullført, 31.03.2022.

### **Frivillig deltakelse**

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn.

Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert.

Dersom du ønsker å delta eller har spørsmål til studien, ta kontakt med undertegnede (se under).

Studien er meldt til Personvernombudet for forskning, Norsk samfunnsvitenskapelig datatjeneste AS.

## **Samtykke til deltakelse i studien**

Jeg har mottatt informasjon om studien, og er villig til å delta

-----  
(Signering, dato)

Jeg samtykker til å delta i intervju

Med vennlig hilsen

Roger Bakken  
Stipendiat  
Høgskolen i Oslo og Akershus  
Fakultet for Lærerutdanning og internasjonalisering  
Institutt for Yrkesfaglærerutdanningen  
Tlf: 95997770  
Mail: rbakken@hioa.no

Leder for MECVET-prosjektet:  
Leif Lahn  
Høgskolen i Oslo og Akershus  
Fakultet for Lærerutdanning og internasjonalisering  
Institutt for Yrkesfaglærerutdanningen  
Mail: leifchristian.lahn@hioa.no

# Godkjenner



Oslo kommune  
**Utdanningsetaten**

Roger Bakken  
Høgskolen i Oslo og Akershus

Date: 06.02.2015

Deres ref:

Vår ref (saknr):

Saksbeh:

Arkivkode:

14/06622-34

Ida Gløsen Josefsen, 23430322

## Fagprøver i elektrikerfaget

Utdanningsetaten viser til forespørsel fra Roger Bakken og forskningsprosjektet:  
Helhetlig vurdering for profesjonslæring i et yrkesdidaktisk perspektiv.

Avdeling for fagopplæring gir Roger Bakken, stipendiat fra Høgskolen i Oslo og Akershus, tilgang til oppgaver, vurderingskriterier og kandidatens besvarelser på fagprøven i elektrikerfaget. Bakken får de nødvendige dokumenter direkte fra prøvenemnda.

Oppgaver, vurderingskriteriene og kandidatens besvarelser skal være anonymisert.

Med hilsen

Marianne Waage  
seksjonssjef

Ida Gløsen Josefsen  
fagkonsulent

**Dokumentet er elektronisk godkjent**

---

Utdanningsetaten

Besøksadresse:  
Strømsveien 102  
0663 OSLO  
Postadresse:  
Pb 6127 Etterstad, 0602 OSLO

Telefon: 02 180      Org.nr.: 976820037  
Telefaks: 22 65 79 71  
postmottak@ude.oslo.kommune.no  
www.ude.oslo.kommune.no



**Fra utdanningsdirektoratet i anvendelse av eksamensoppgave**

**From:** Ivar Svensson <ivar.svensson@udir.no>

**Sent:** Tuesday, September 17, 2019 11:29 AM

**To:** Roger Bakken <rbakken@oslomet.no>

**Subject:** SV: Godkjennelse

Du gis herved tillatelse til å bruke forberedelses- og eksamensdel av ELE3002 elektrikerfaget, gitt våren 2015. Dette anses som en forlengelse av tidligere tillatelse til å følge prosessen med utarbeidelse av eksamener i ELE3002 elektrikerfaget.

Hilsen IvarS