

Trusler for validitet i vurdering i biologi 2

En kasusstudie om vurderingspraksis i
biologi 2 og trusler for validitet

Susanne Dyrstad

Master i realfag

Innlevert: juni 2018

Hovedveileder: Alex Strømme, PLU

Medveileder: Lise Sandvik, PLU

Norges teknisk-naturvitenskapelige universitet
Institutt for lærerutdanning

Forord

Denne masteroppgaven i biologididaktikk markerer slutten på lektorutdanningen i realfag ved NTNU i Trondheim. Jeg ser tilbake på et intenst og krevende arbeid, som samtidig har vært svært interessant og lærerikt.

Først vil jeg rette en stor takk til min hovedveileder, Alex Strømme, for alle gode råd og konstruktive veiledninger. Ditt faglige engasjement inspirerer. I tillegg vil jeg takke Lise Vikan Sandvik og Henning Fjørtoft for gode innspill underveis. Jeg vil også takke biologilæreren og biologielevene som gjorde denne studien mulig.

Avslutningsvis vil jeg takke alle støttespillerne som har vært uvurderlig gode å ha under hele studieperioden. Marius, du er enestående! Mamma og pappa, jeg hadde ikke klart dette uten dere! Venner og medstudenter, dere har gjort studietiden riktig fin!

Tusen hjertelig takk alle sammen.

Trondheim, juni 2018

Susanne Dyrstad

Sammendrag

Hensikten med denne studien har vært å identifisere trusler for validitet som biologilærere kan møte på i ulike faser av vurderingspraksisen, samt hvordan disse truslene kan identifiseres. Studien kan beskrives som en kasestudie der kvalitative metoder er brukt for å besvare problemstillingen. En trussel for validitet kan være hvis en prøve krever andre kognitive ferdigheter hos elevene enn kompetansemålene. For å undersøke hvilke kognitive ferdigheter oppgaver gitt på en prøve i genetik i biologi 2 krevde av elevene, ble prøveoppgavene analysert og sammenlignet med hvilke kognitive ferdigheter kompetansemålene krevde av elevene. Analyseverktøyet var basert på Blooms taksonomi og Webbs nivåinndeling. Funnene indikerte at både kompetansemålene og prøveoppgavene for det meste krevde lavere- og middels kognitive ferdigheter hos elevene.

For å sette denne vurderingssituasjonen i en større kontekst, ble biologilæreren og fem biologielever intervjuet om vurderingspraksisen i biologi 2-lassen. Fokusgruppeintervjuet med elevene ble analysert ved hjelp av tematisk koding, mens intervjuet med læreren ble brukt som sekundær empiri. Analysen av elevenes refleksjoner resulterte i fire analytiske påstander som indikerer trusler for validitet i vurderingspraksisen: *uklare læringsmål, nivåforskjell på prøve og eksamen, bruk av lærebok og liten bruk av tilbakemeldinger.*

På bakgrunn av analysen av kompetansemålene og prøveoppgavene, samt elevenes refleksjoner rundt vurderingspraksisen, har jeg supplert validitetskjeden med spørsmål som biologilærere kan bruke som et verktøy for å identifisere trusler for validitet i vurderingspraksisen.

Abstract

The aim of the study has been to identify threats of validity that biology teachers may face during different phases of the assessment process, as well as how these threats may be identified. The study is described as a case study, where qualitative methods has been used for answering the research problem. A threat to validity is when a test requires other cognitive skills than presented in the competence aims. In order to investigate cognitive skills, a test from a biology 2 class about genetics was analyzed. These analyses were based on the competence aims of the subject and the cognitive skills to each of them. Both Blooms taxonomy and Webbs level of classification, was used. The findings indicate that both the competence aims as well as the test, required lower- and medium cognitive skills.

As well as the analyses, the teacher and five students from the same class presented, were interviewed. A focus group interview was analyzed with thematically coding, and the interview with the teacher was used as secondary empiricism. The analyses from the student interview gave four analytical allegations which indicate a threat for validity in assessments: *unclear learning objectives, level differences in tests and examinations, use of textbook and little use of feedback.*

In context of the analyzes of the competence aims and tests, as well as reflections from the students, the validity chain has been supplemented with questions that biology teachers may use as tools for identifying validity threats in assessment situations.

Innholdsfortegnelse

1	INNLEDNING	1
1.1	BAKGRUNN	1
1.2	PROBLEMSTILLING OG FORSKNINGSSPØRSMÅL	2
1.3	BEGREPSAVKLARINGER	2
1.3.1	<i>Vurderingspraksis</i>	3
1.3.2	<i>Validitet</i>	3
1.3.3	<i>Validitetskjede</i>	4
1.3.4	<i>Trusler for validitet</i>	4
1.4	OPPBYGGING AV STUDIEN	4
2	TEORI	5
2.1	VURDERING	5
2.1.1	<i>Vurdering – et kontekstuel element i elevenes læringsomgivelser</i>	7
2.1.2	<i>Validitet</i>	7
2.1.3	<i>Modell for analyse og utvikling – validitetskjede</i>	8
2.1.4	<i>Trusler for validitet i vurdering</i>	10
2.1.5	<i>Validitet i klasseromskontekst</i>	15
2.1.6	<i>Forskning på vurderingspraksis og validitet i norsk skole</i>	17
2.2	VURDERING AV ELEVENES KOMPETANSE I BIOLOGI	17
2.2.1	<i>Hva kjennetegner biologi som skolefag?</i>	17
2.2.2	<i>Kompetanse i biologi</i>	17
2.2.3	<i>Hvordan kan elevenes kompetanse i biologi vurderes?</i>	18
2.2.4	<i>Skriveoppgaver og flervalgsoppgaver</i>	18
2.2.5	<i>Bruk av lærebok i undervisningen</i>	19
2.2.6	<i>Er eksamen styrende?</i>	19
2.3	KOGNITIVE FERDIGHETER	19
2.3.1	<i>Blooms taksonomi</i>	20
2.3.2	<i>Webbs nivåinndeling</i>	21
3	METODISK TILNÆRMING	23
3.1	KONTEKST	23
3.1.1	<i>Metodisk rammeverk</i>	24
3.2	FORSKNINGSDESIGN	24
3.3	DATAINNSAMLING	24
3.4	ANALYSEVERKTØY OG ANALYSETRINN	26
3.4.1	<i>Analyseverktøy og analysetrinn i innholdsanalysen</i>	26
3.4.2	<i>Analysemetode og analysetrinn i fokusgruppeintervjuet</i>	27

3.5	METODEKVALITET.....	28
3.5.1	<i>Kredibilitet</i>	28
3.5.2	<i>Overførbarhet</i>	29
3.5.3	<i>Avhengighet</i>	29
3.5.4	<i>Bekreftbarhet</i>	29
3.6	ETISKE BETRAKTNINGER	30
4	ANALYSE	33
4.1	KOMPETANSEMÅLENE INDIKERER BRUK AV LAVERE OG MIDDELS KOGNITIVE FERDIGHETER	33
4.2	PRØVEOPPGAVERNE INDIKERER BRUK AV LAVERE OG MIDDELS KOGNITIVE FERDIGHETER.....	35
4.2.1	<i>Eksempler på oppgaver som er plassert innen kategorien lav</i>	36
4.2.2	<i>Eksempler på oppgaver som er plassert innen kategorien middels</i>	37
4.3	BIOLOGIELEVENES REFLEKSJONER.....	39
4.3.1	<i>Kompetansemål</i>	39
4.3.2	<i>Vurderingssituasjon</i>	40
4.3.3	<i>Vurdering</i>	42
4.4	OPPSUMMERING AV FUNN	44
5	DISKUSJON.....	45
5.1	HVORDAN KAN KOMPETANSEMÅL OG PRØVEOPPGAVER ANALYSERES FOR Å GI INNSIKT I HVILKE KOGNITIVE FERDIGHETER DE KREVER HOS ELEVENE?	45
5.1.1	<i>Samsvar mellom kompetansemål og prøveoppgaver</i>	45
5.1.2	<i>Styrker med analyseverktøyet</i>	46
5.1.3	<i>Svakheter med analyseverktøyet</i>	46
5.2	HVORDAN KAN EN ANALYSE AV VALIDITETSKJEDEN BRUKES SOM ET VERKTØY AV LÆREREN FOR Å SIKRE BEST MULIG VALIDITET I VURDERINGSPRAKSISEN?.....	47
5.2.1	<i>Kompetansemål</i>	48
5.2.2	<i>Vurderingssituasjon</i>	49
5.2.3	<i>Vurdering</i>	50
5.3	VALIDITETSKJEDEN MED SUPPLERENDE SPØRSMÅL.....	51
5.4	DIDAKTISKE REFLEKSJONER	53
5.5	STUDIENS KVALITET.....	54
6	AVSLUTNING	57
6.1	AVSLUTTENDE REFLEKSJONER.....	57
6.2	SVAR PÅ FORSKNINGSSPØRSMÅL.....	57
6.3	KONKLUSJON	58
6.4	VEIEN VIDERE.....	59
7	LITTERATURLISTE.....	61

Oversikt over vedlegg

Vedlegg A: Prøve i genetikk i biologi 2

Vedlegg B: Intervjuguide biologielever

Vedlegg C: Intervjuguide biologilærer

Vedlegg D: Bekreftelse på NSD-søknad

Vedlegg E: Samtykkeskriv

Figurliste

Figur 1. Validitetskjede (Crooks, Kane & Cohen, 1996)	s. 9
Figur 2. Validitetskjede (Sandvik & Fjørtoft, 2016)	s. 14
Figur 3. Flytskjema som viser analysekategorier fra elevenes refleksjoner	s. 28
Figur 4. Figur over kognitive nivå på prøven.	s. 35

Tabelliste

Tabell 1. Blooms taksonomi	s. 20
Tabell 2. Webbs nivåinndeling	s. 22
Tabell 3. Analyseverktøy (Blooms taksonomi og Webbs nivåinndeling)	s. 26
Tabell 4: Trinn i validitetskjede med supplerende spørsmål	s. 52

1 Innledning

1.1 Bakgrunn

Realfagsprøver gitt i den norske skolen har vist seg å bære større preg av gjengivelse enn det å forstå. Samtidig er samfunnet elevene skal delta i preget av store og raske endringer innen teknologi, globalisering, klima og generell kunnskapsutvikling (Bergem et al., 2015).

Momsen et al. (2013) poengterer at elever vil ha fokus på å lære seg det de må kunne til de ulike prøvene de har i løpet av skoleåret. Med andre ord vil valg som læreren gjør knyttet til vurderingen være et av flere viktige kontekstuelle element i elevenes læringsomgivelser.

I løpet av de siste tiårene har det blitt større oppmerksomhet omkring kvaliteten i det norske utdanningssystemet, noe som igjen har medført et søkelys på hvilken rolle vurdering har i elevenes læringsprosess og validiteten i denne vurderingen. Ifølge William (2011) er vurdering blant de mest sentrale aspektene ved god undervisning, og validitet regnes av flere som den viktigste kvaliteten for vurderinger (Crooks, Kane & Cohen, 1996; Sandvik & Fjørtoft, 2016; Stobart, 2012). Dette viser viktigheten av å forske på dette feltet.

Forskning som er gjort i den norske skolen viser at det finns mange ulike vurderingspraksiser, og at det i mange tilfeller oppdages ulike trusler for validitet i vurderingene (Sandvik et al, 2014). Eksempelvis at elever ikke bruker tilbakemeldinger som gis i vurderingssituasjoner, eller at elevene ikke er klar over hva de blir vurdert etter. På bakgrunn av blant annet disse utfordringene har det blitt igangsatt satsninger som skolebasert kompetanseutvikling i vurdering – SKUV. Her er det forskergrupper som kartlegger de ulike skolens behov før aktiviteter igangsettes med formål om å utvikle vurderingskompetansen.

Som kommende lærer i biologi har jeg en motivasjon til å fordype meg i vurdering, da jeg gjennom lektorstudiet har savnet mer kunnskaper innen dette feltet. I løpet av praksisperiodene har jeg selv fått utfordret vurderingsferdighetene, og sett nødvendigheten av å ha ulike verktøy som kan bidra til å sikre kvaliteten i vurderingen. Eksempelvis det å operasjonalisere kompetansemål, designe gode vurderingsoppgaver og bruke vurdering til å bidra positivt i elevenes læringsprosess. Dette, i tillegg til at jeg ikke har funnet forskning som

er gjort på vurderingspraksiser i biologi i den norske videregående skolen, bidro til at jeg ønsket å få en dypere innsikt innen dette feltet.

1.2 Problemstilling og forskningsspørsmål

Formålet med studien er å få kunnskaper om trusler for validitet som biologilærere kan møte på i ulike faser av vurderingspraksisen, samt hvordan truslene kan identifiseres. Gjennom analyse av kompetansemål og oppgaver gitt i en skriftlig prøve i genetik, samt intervju med biologielever om vurderingspraksisen vil jeg få en større innsikt i faktorer som påvirker validiteten i vurderingspraksisen. I tillegg ble biologilæreren intervjuet, og dette brukes som sekundær empiri. Studien har ikke til hensikt å kritisere biologilærerens praksis, men å få en dypere innsikt i vurderingspraksisen. På bakgrunn av empiri og teori vil jeg bidra med å utvikle et sett med spørsmål som biologilærere kan bruke for å undersøke validiteten i vurderingspraksisen og identifisere trusler for validitet.

Med bakgrunn i dette har jeg valgt å se på følgende problemstilling:

Hvilke trusler for validitet kan biologilærere møte på i ulike faser av sin vurderingspraksis, og hvordan kan disse identifiseres?

For å kunne svare på problemstillingen vil jeg svare på følgende forskningsspørsmål:

1. Hvordan kan kompetansemål og prøveoppgaver analyseres for å gi innsikt i hvilke kognitive ferdigheter de krever av elevene?
2. Hvordan kan validitetskjeden og identifisering av trusler for validitet brukes som verktøy av biologilærere for å sikre best mulig validitet i vurderingspraksisen?

1.3 Begrepsavklaringer

I denne seksjonen vil jeg gi en beskrivelse av begrepene *vurderingspraksis*, *validitet*, *validitetskjede* og *trusler for validitet*, som er sentrale for studien.

1.3.1 Vurderingspraksis

Ifølge den amerikanske rapporten *Knowing what Students Know* (Pellegrino, Chudowsky & Glaser, 2001) er dette beskrivende for hva en *vurderingspraksis* innebærer:

- En teoretisk modell for å ha en forståelse av hvordan elevkunnskap kan komme til uttrykk ved hjelp av ulike representasjonsformer, i tillegg til å forstå hvordan elevene utvikler kompetanse innenfor et gitt emne.
- Vurderingssituasjoner eller oppgaver som muliggjør å observere elevenes prestasjoner
- En fortolkningsmetode der læreren bruker observasjoner av elevenes prestasjoner til å trekke slutninger

Lærere må kunne bruke et vidt spekter av fortolkningsmetoder for å kunne si noe om elevenes læring, da ulike typer oppgaver og vurderingssituasjoner vil kunne synliggjøre forskjellige observasjoner og prestasjoner. Eksempelvis må læreren kunne bruke elevenes skriftlige arbeid som utgangspunkt for tolkninger av bevis på læring. Disse tolkningene baseres på fagkompetanse, klasseromserfaring og fagdidaktiske forståelse. I praksis må da læreren kunne se sammenheng mellom læreplanmål, undervisning og vurdering hvis ikke kan vurderingen oppleves som svekket (Sandvik et al., 2012).

1.3.2 Validitet

En vanlig definisjon av validitet er «an evaluative summary of both evidence for and the actual – as well as potensial – consequence of score interpretation and use.» (Messick, 1995, s. 742). Validiteten sier noe om kvaliteten i vurderingen. Det vil si relevansen av bevisene lærerne samler inn om elevenes læring, tolkning av disse bevisene i forhold til formålet med vurderingen og konsekvensene av vurderingen. Med andre ord, når læreren skal undersøke validiteten i vurderingspraksisen, bør han/hun evaluere om tolkningene av elevenes prestasjoner samsvarer med kompetansemålene. Læreren bør også stille spørsmål ved om bevisene for elevenes kompetanse gir et realistisk bilde av elevens kunnskaper innenfor et gitt område av læreplanen. Dette i tillegg til å evaluere om konsekvensene av vurderingen samsvarer med elevens prestasjon og elevens behov videre i undervisningen (Kane, 2016; Messick, 1995).

1.3.3 Validitetskjede

En validitetskjede representerer prosessen fra planlegging av vurdering, via å lage vurderingsoppgaver til konsekvensene av vurderingen (Crooks, Kane & Cohen, 1996). I kapittel 2.1.3. gis det en inngående beskrivelse.

1.3.4 Trusler for validitet

Messick (1994) argumenterer for at evaluering av tilsiktede og utilsiktede konsekvenser av prøver er avgjørende for validiteten av fortolkninger og bruk av dem. Det er da hensiktsmessig å identifisere mulige aspekter som kan bidra til å redusere validiteten til vurderingen. Innenfor validitetsforskning brukes begrepet *trusler* om faktorer som kan redusere validiteten (Crooks, Kane & Cohen, 1996; Messick 1987). Med andre ord, når en skal evaluere validiteten i vurderingen må man undersøke hvilke mulige faktorer som kan utgjøre en trussel for validiteten.

1.4 Oppbygging av studien

Studien består av i alt seks kapitler. Etter dette introduksjonskapitlet følger kapittel 2 som tar for seg studiens teoretiske forankring, etterfulgt av kapittel 3 hvor metodisk tilnærming presenteres. Samlet danner disse to kapitlene grunnlaget for å tolke og analysere studiens empiri. I kapittel 4 legger jeg frem studiens funn, før disse diskuteres i kapittel 5 hvor også studiens kvalitet drøftes. Avslutningsvis, i kapittel 6, kommer jeg med en avsluttende refleksjon, svar på forskningsspørsmålene, konklusjon ut ifra problemstillingen og veien videre.

2 Teori

I dette kapitlet presenterer jeg den teoretiske forankringen som danner grunnlaget for analysen, og videre drøfting av problemstilling. Først vil jeg gå nærmere inn på vurdering generelt, og deretter validitetskjeden og trusler for validitet i vurderingen spesielt. Deretter belyses vurdering i biologi med vekt på hva kompetanse i faget innebærer og hvordan den kan vurderes. I siste delkapittel tar for en oppdeling av kognitive ferdigheter basert på Blooms taksonomi og Webbs nivåinndeling.

2.1 Vurdering

Ifølge William (2011) er vurdering blant de mest sentrale aspektene ved god undervisning. I boken *Vurdering og læring i en elevaktiv skole* presenterer Dale og Wærness (2006) hensiktene med vurdering, som for eksempel krav til seleksjon til videre studier og arbeidsliv og sertifisering til bestemte kompetansenivåer. De vektlegger også et annet aspekt ved vurdering, nemlig at den bør ha en intensjon om å identifisere elevens læring for å planlegge det videre læringsarbeidet.

Det er vanlig å skille mellom formativ og summativ vurdering i vurderingsforskning. Ved summativ vurdering er hensikten først og fremst å kartlegge elevenes nåværende kompetanse (Black, Harrison, Lee, Marshall & Dylan, 2003). Formativ vurdering omhandler vurderingsformer der formålet er å skape positive forandringer i den videre læringsprosessen (Engh, 2011). For at en lærer skal mestre formativ vurdering, må han/hun kunne finne ut hvor eleven er i læringen, kunne lage læringsmål og finne ut hva som kreves for å nå de (Sadler, 1989). Med andre ord kan krav til seleksjon og sertifisering være eksempler på summativ vurdering i form av eksamenskarakterer og standpunkt karakterer. Det formative aspektet ved vurdering skal ivareta en viktig hensikt med vurdering, nemlig læring. Jeg har til nå fremstilt et skille mellom disse to retningene, men denne forståelsen har blitt utfordret.

Ifølge Black og William (1998, 2009) må ikke nødvendigvis formativ og summativ vurdering betraktes som selvstendige kategorier, da alle vurderingspraksiser kan ha formative sider dersom de benyttes til planlegging av undervisningen videre i læringsforløpet. De sier at formativ vurdering beskrives som alle aktiviteter der lærer innhenter informasjon som brukes

som tilbakemelding for å tilpasse videre læringsforløp, i tillegg til elever som vurderer seg selv. Poenget er at vurdering får en formativ karakter når informasjonen brukes til å tilpasse undervisningen etter elevenes behov. Med andre ord er det avgjørende om informasjonen fra vurderingssituasjonen brukes, og om den støtter læringsprosessen til eleven (William, 2011). Taras (2005) beskrivelse av formativ vurdering er at den bygger på den summative vurderingen i en prosess som skal støtte læring, og hvor tilbakemeldingene er en viktig del av vurderingsprosessen.

I Black og Williams (2009) studier trekker de frem at formativ kriteriebasert vurdering fører til bedre læring. Det vil si å vurdere elevprestasjoner opp mot forhåndsbestemte læringsmål og kriterier, enten det innebærer at læreren bruker disse aktivt når tilbakemeldinger gis til elevene eller at elevene bruker de når de vurderer eget arbeid eller medelevers arbeid. Når elever og lærer har en felles forståelse for grunnlaget for vurderingen, kan det kalles et tolkningsfellesskap hvor læringsprosessen blir mer transparent.

Når man ser formativ vurdering i et dynamisk perspektiv, kalles det gjerne *vurdering for læring*. Et viktig prinsipp er at elevene er aktive i sin egen læringsprosess, som kan innebære å delta i vurderingsarbeid, felles forståelse for læringsmålene, få læringsfremmende tilbakemeldinger og planlegge undervisningen med utgangspunkt i læringsmålene (Black & William, 1998, 2009). Lærerens rolle blir da å lage og implementere et godt læringsmiljø, mens elevene har ansvaret for læringen (Black & William, 2009). Ifølge William og Thomson (2007) kan formativ vurdering oppsummeres slik:

- Tydelige mål og vurderingskriterier
- Effektive klasseromsdiskusjoner og gode vurderingsoppgaver som dokumenterer elevenes kompetanse
- Tilbakemelding som fremmer læring
- Elevene brukes som ressurser for hverandres læring
- Elevenes evne til å lære å lære

I denne studien legges det særlig vekt på de tre første punktene som gjelder *tydelige mål, vurderingsoppgaver og tilbakemeldinger*. Prøven som analyseres plasseres innenfor

perspektivet om at formative og summativ vurdering ikke er atskilte, da den ble brukt til både å gi tilbakemelding på elevenes kompetanse på tidspunktet, tilbakemeldinger for hva som kunne forbedres fremover og som bevis for elevenes kompetanse. Med andre ord kan prøven omtales som en summativ prøve som også brukes formativt.

2.1.1 Vurdering – et kontekstuel element i elevenes læringsomgivelser

Vurderinger, både formative og summative, representerer et kontekstuel element i læringsomgivelsene til elevene. Dette elementet kan forme elevenes læring gjennom deres oppfatninger av innhold og kognitive ferdigheter som trengs for å gjøre det bra (Hall, Watkins, Coffey, Cooke & Redish. 2011). Ifølge Entwistle og Entwistle (1992) kan elevene selektivt eller strategisk velge å studere det som best mulig sammenfaller med deres forventninger av vurderingene. Eksempelvis vil elevene kunne velge en overflatelæring hvis de møter prøver med flervalgsoppgaver som kun tester faktakunnskaper (Momsen et al., 2013).

Videre vil det være interessant å se på validiteten i den vurderingen som gjøres i klasserommet.

2.1.2 Validitet

Bonnier (2013) sier at vurderinger i klasserommet noen ganger inkluderer målinger, men alltid vil omfatte innhenting av informasjon om elevenes ferdigheter, tolking av denne informasjonen og å bruke den videre i et læringsformål. Det har blitt satt søkelys på at denne tolkningen og bruken av informasjonen til et formål bør valideres. Validitet er et begrep som er sentralt i denne studien, og regnes av flere som den viktigste kvaliteten for vurderinger (Crooks, Kane & Cohen, 1996; Sandvik & Fjørtoft, 2016; Stobart, 2012). Klassisk testteori tar utgangspunkt i at validitet er en egenskap ved en prøve. Prøven er valid hvis den måler det den oppgir at den skal måle (Stobart, 2009). Validitet har flere aspekter ved seg, for eksempel at det er interferensen fra testresultatene som har betydning for validiteten eller at begrepet må forstås i en større kontekst (Stobart, 2009). Gjennom tiden har forskere skilt mellom ulike typer validitet. *Innholdsvaliditet* handler om innholdet i en prøve representerer et bestemt faglig innhold, for eksempel om en prøve gjenspeiler kompetansemålene som står i læreplanen. *Kriteriebasert validitet* tar for seg gyldigheten av slutningene en lærer kan ta som

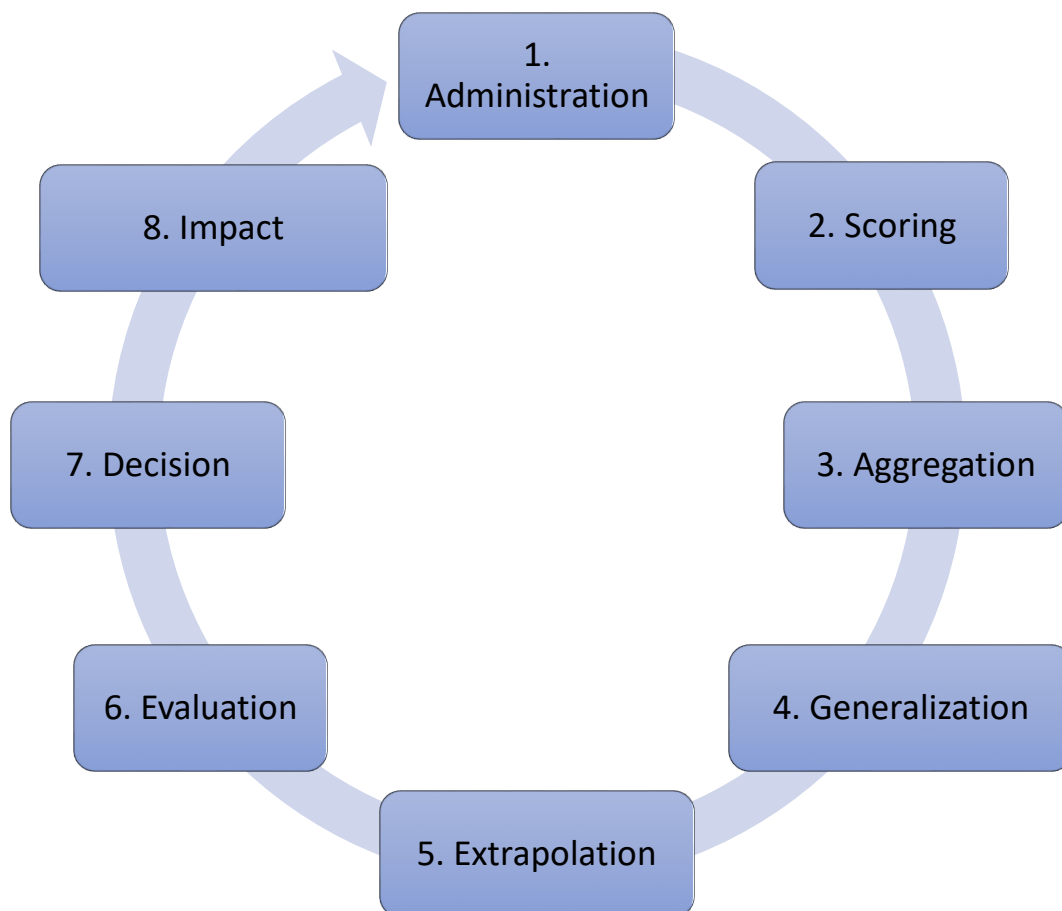
følge at prøveresultatene hos en elev basert på kriterier. Den tredje retningen, *konstruktvaliditet*, omhandler forståelser og relasjoner som samlet medfører at det kan trekkes slutninger til en elevs kompetanse på et gitt område (Sandvik et al. 2014). Messick (1995) er en av dem som hevder at validitet ikke kan deles opp slik, og at begrepet må ses i en større kontekst. Ifølge han vil da vurdering handle om fortolkninger og slutninger, samt at konsekvensvaliditet må inkluderes med tanke på konsekvenser ulike prøver har for elevens læring.

2.1.3 Modell for analyse og utvikling – validitetskjede

Det formelle læreplandokumentet som utdanningssystemet skal forholde seg til er en av de viktigste styringsfaktorene i skolen (Imsen, 2016). Den gjeldende læreplanen i biologi 2 skal utgjøre en ramme for biologilærerens arbeid, og er et dokument som læreren må tilpasse den lokale konteksten. Læreren må gjøre tolkninger av læreplanen, slik at den kan realiseres i klasserommet. Det innebærer eksempelvis bestemmelse av læringsmål, arbeidsmåter og vurderingssituasjoner. Goodlad (1979) vektlegger fem ulike dimensjoner av læreplaner. *Ideenes læreplan* er de tanker som finnes om hvordan en læreplan bør være med tanke på form og innhold, og den vil påvirkes av næringsliv, politikk og samfunnsutvikling. *Den formelle læreplanen* er selve læreplandokumentet som fungerer som ramme for skolens virksomhet. *Den oppfattede læreplanen* er de tolkninger lærere og andre gjør når de leser den formelle læreplanen, og legger dermed grunnlag for eksempelvis planlegging av undervisning. Med andre ord vil den oppfattede læreplanen kunne variere fra person til person. Den operasjonaliserte læreplanen er det som faktisk skjer i undervisningen, altså utførelsen av den oppfattede læreplanen. *Den erfarte læreplanen* inngår i elevenes oppfattelse av undervisningen, samt foreldres og samfunnets oppfattelse av skolens virksomhet. Med andre ord er det mange tolkningsledd fra den opprinnelige intensjon til elevenes opplevelser av lærerens operasjonalisering av læreplanen. I denne studien vil jeg se på aspekter ved den *formelle læreplanen*, *den oppfattede læreplanen* og *den erfarte læreplanen*.

Lærernes fortolkningsprosesser fra kompetansemål til omsetting av dem i undervisningen, viser viktigheten av å undersøke validiteten i den lange kjeden av fortolkninger, og ikke bare om prøveresultatene regnes som gyldige (Sandvik & Fjørtoft, 2016). I arbeidet med den oppfattede læreplanen er det viktig at læreren kan samle inn gode bevis for elevenes læring, slik at han/hun kan fortolke det og tilrettelegge for elevens utvikling. En lærer som ikke har

denne kompetansen vil risikere at læringsmålene ikke blir realisert på en best mulig måte for elevene (Sandvik & Fjørtoft, 2016). *Validitetskjeden* presentert i Figur 1 representerer prosessen fra planlegging av vurdering, via å lage vurderingsoppgaver til konsekvensene av vurderingen (Crooks, Kane & Cohen, 1996; Sandvik & Fjørtoft, 2016). Den er bygget opp av åtte trinn der hvert trinn tar utgangspunkt i ulike valideringskriterier. Hensikten med modellen er å gi den som skal validere vurderingen flere trinn å følge for å sikre og evaluere validiteten (Crooks, Kane & Cohen, 1996). Denne modellen tar utgangspunkt i standardiserte tester.



Figur 1. Validitetskjeden utviklet av Crooks, Kane og Cohen (1996), s. 268.

Det første trinnet tar for seg *administrering* (administration) av vurderingen, og her kan mulige trusler og lærerens bestemmelser påvirke elevenes prestasjoner. Det andre trinnet er *poenggiving* (scoring). Det tredje trinnet er *aggregering* (aggregation) av poenggiving. Fjerde trinn er *generalisering* (generalization) av elevenes kunnskaper i kompetansen som er

prøvd ut. Femte trinn er *ekstrapolering* (extrapolation) av kompetansen som læreren har tatt utgangspunkt i å teste, for eksempel at det settes en karakter. Det sjette trinnet er *evaluering* (evaluation) av vurderingsprosessen i forhold til intensjonene, og dette vil påvirke lærerens vurdering av elevenes prestasjoner. Det syvende trinnet er forbeholdt lærerens *avgjørelser* (decision) på bakgrunn av informasjonen fra vurderingen, og vil kunne ha innvirkning på hva som gjøres videre i undervisningen. Siste trinn i validitetskjeden er *konsekvensene* (impact) av vurderingen, og tar for seg hvordan hele vurderingsprosessen kan påvirke eleven (Crooks, Kane & Cohen, 1996).

2.1.4 Trusler for validitet i vurdering

Ifølge Crooks, Kane og Cohen (1996) vil det første steget i valideringsprosessen være å evaluere hva som er viktig i kjeden med tanke på formålet til vurderingen, og de sier eksplisitt at validering kun kan skje når det intenderte formålet med vurderingen er forstått.

Eksempelvis vil det ikke være hensiktsmessig å se på trusler relatert til aggregering av poeng på tester bestående av få oppgaver. I tillegg vektlegger de at den svakeste delen i kjeden må identifiseres, da validiteten ikke er sterkere enn det svakeste trinnet. I de neste avsnittene vil jeg gå nærmere inn på hvert trinn i validitetskjeden, og beskrive mulige trusler for validitet.

1. Administrering

Når man undersøker validiteten ved administrering av vurderingsoppgaver vil det være hensiktsmessig å se på konteksten og omstendighetene rundt vurderingen. Faktorer som kan nevnes her er angst, motivasjon, vurderingsbetingelser og samsvar mellom oppgaver og svar (Crooks, Kane & Cohen, 1996). For eksempel kan forebygging av angst hos elevene øke validiteten, men samtidig er det viktig at læreren tar hensyn til de andre faktorene som kan spille inn på validiteten.

2. Poenggiving

Crooks, Kane og Cohen (1996) vektlegger fem trusler når det gjelder poenggiving av elevenes prestasjoner. Den første trusselen gjelder situasjoner der deler av prestasjonen til eleven blir utelatt fra vurderingen, for eksempel som følge av feil i rubrikker som brukes til poenggivingen. Kanskje det kun er deler av viktige kvaliteter som beskriver presentasjonen

som blir vurdert? Den andre trusselen tar for seg situasjoner der bestemte typer kriterier eller svar vektlegges i poenggivingen, noe som kan beskrives som konstrukt underrepresentasjon (Messick, 1995). Den tredje trusselen handler om inkonsistent vurdering, eksempelvis at læreren ikke er konsistent i sin egen vurdering. Den fjerde trusselen fokuserer på at poenggivingen kan være for analytisk, det vil si at læreren retter for stort fokus på poengsetting av hver del av prøven og mister det store bildet. Den femte trusselen derimot, belyser tilfeller der rettingen er for holistisk. Det kan være at elevene ikke får tilbakemelding på prestasjonene på de ulike delene av vurderingen, eksempelvis når de bare sitter igjen med en karakter uten noen annen informasjon om deres prestasjoner (Crooks, Kane & Cohen, 1996).

3. Aggregering

Dette trinnet tar for seg når poeng fra hver oppgave settes sammen til en total poengsum, og her er det to trusler som identifiseres. Validiteten til vurderingen kan svekkes hvis oppgavene er for sprikende, eksempelvis hvis prøven inneholder flere forskjellige temaer. Den andre trusselen er når læreren vektlegger ulike aspekter elevenes prestasjoner.

4. Generalisering

Det fjerde trinnet i validitetskjeden omhandler hvordan oppgaver brukes som bevis for elevenes kunnskaper innen et spesifikt emne. For eksempel kan liten tid til å fullføre prøven, ulike oppgaveformater og tilnærminger kunne være trusler for validiteten. Andre trusler kan være hvis vurderingen inneholder for få oppgaver til at elevene får vist sin kompetanse eller trusler knyttet til kriterier og poenggivning.

5. Ekstrapolering

Når læreren skal gjøre en vurdering av elevenes kompetanse, vil gjerne prøven bestå av ulike delkompetanser der noe er utelatt. Læreren må derfor vurdere om konstruktene som skal vurderes er tilstrekkelig representert. Trusler som om oppgavene representerer det de er tiltenkt fra pensum, og er noen deler av konstruktene utelatt eller gitt for liten oppmerksomhet? Et annet punkt som kan si noe om validiteten, er hvis hele prøven inneholder flervalgsoppgaver. Ville da eleven kunne vist samme resultat innen et gitt emne

om prøven inneholdt skriveoppgaver? Et annet eksempel er hvis læreren har laget oppgaver for å dekke et gitt emne, men ikke oppnår det tiltenkte formålet med tanke på innhold, kvalitet eller kognitiv kompleksitet (Crooks, Kane & Cohen, 1996).

6.Evaluering

Når det gjelder dette trinnet, evaluering, kan en stille spørsmålet om hva poengene som er gitt betyr? Dette innebærer validering av tolkningen av den vurderte prestasjonen og tolkning av konstrukt. Hva sier egentlig elevens prestasjon på denne prøven og han/hennes kompetanse i biologi? Crooks, Kane og Cohen (1996) nevner tre trusler her, nemlig dårlig forståelse av informasjonen som er hentet ut vurderingen, konsekvenser relatert til dårlig forståelse av konstrukt og partisk tolkning av prestasjonen.

7.Avgjørelser

Dette trinnet belyser lærerens avgjørelser som blir tatt på bakgrunn av vurderingen, altså hva gjør han/hun med informasjonen som er samlet inn fra vurderingssituasjonen. Dette kan være om elever trenger spesiell tilrettelegging, eller om det er noe som må gjennomgås i klassen. Ifølge Crooks, Kane og Cohen (1996) vil gode beslutninger resultere i positive konsekvenser for elevene. Et eksempel kan være at læreren ser hva ulike elever trenger mer trening i, og følger opp dette i videre undervisning. Eksempler på trusler her er upassende vurderingsstandarder laget av læreren, eller svekkede pedagogiske beslutninger på hva som skjer videre etter vurderingssituasjonen. Et eksempel på sistnevnte er hvis eleven har prestert bra samlet sett på prøven, men læreren velger å fokusere kun på det som ikke var bra når tilbakemeldinger gis. En konsekvens kan da være at eleven får et misvisende bilde av prestasjonene på prøven, og det kan påvirke motivasjonen videre.

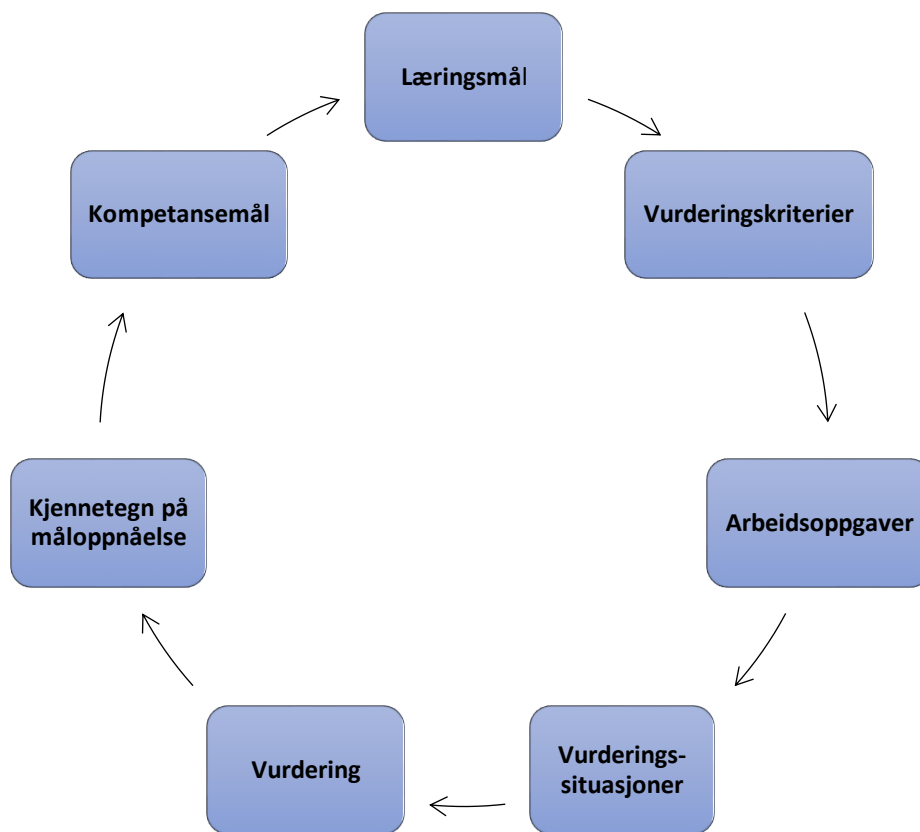
8.Konsekvenser

Det siste trinnet i validitetsskjeden er hvilke konsekvenser vurderingen får for elevene og andre involverte. I tillegg til de direkte og indirekte konsekvensene av vurderingen, er det også viktig å ta effekten av opplevelsen av hele vurderingsprosessen i betraktning. Gis det formativ tilbakemelding som kan bidra til å utvide elevenes kunnskaper på feltet? Andre positive konsekvenser som nevnes er økt motivasjon, større tro på egne ferdigheter og

fremtidig prestasjon. En trussel kan være hvis slik tilbakemelding ikke gis, i tillegg til negative konsekvenser som nedsatt motivasjon eller økt angst. En annen konsekvens som trekkes frem er fokus på faktakunnskaper på bekostning av høyere kognitive ferdigheter.

Denne åtteledd-kjedede modellen som Crooks, Kane og Cohen (1996) har utviklet tar utgangspunkt i standardiserte tester, men er også grunnlaget for en annen validitetskjede som tar utgangspunkt i lærerens arbeid med læreplanen, utvikling av vurderingsoppgaver og tolkning av elevsvar. Figur 1 fra Crooks, Kane og Cohens (1996) arbeid tar utgangspunkt i standardiserte tester, mens validitetskjeden slik den er presentert i Figur 2 har en mer fleksibel bruk og er mer tenkt som et forslag enn en streng prosedyre¹. Validitetskjeden belyser vurdering som et prosessuelt fenomen framfor en enkelthendelse i skolehverdagen (Sandvik & Fjørtoft, 2016), og kan dermed være til hjelp for å undersøke hele vurderingsprosesser fra arbeid med læreplan til tolkning av ulike elevsvar.

¹ Personlig meddelelse via epost med Henning Fjørtoft 4. juni 2018 på spørsmål om hvordan deres modell ble utviklet med utgangspunkt i modellen til Crooks, Kane og Cohen (1996).



Figur 2. Validitetskjeden presentert av Sandvik og Fjørtoft (2016), s. 49.

Sanvik og Fjørtoft (2016) nevner ulike valideringsprosesser som kan styrke validiteten av vurderingsarbeidet. Eksempelvis samarbeid mellom lærere som utvikler prøver og når lærere og elever utvikler vurderingskriterier sammen. Begge eksemplene kan beskrives som tolkningsfellesskap som kan bidra til å øke validiteten i vurderinger. Bonner (2013) legger vekt på at bevis på læring i klasserommet må baseres på oppgaver som er tilpasset undervisningen som har foregått i forkant, det vil si at det er en tydelig sammenheng mellom undervisning og vurdering også kan bidra til å styrke validiteten i vurderingen.

Hvis man skal undersøke alle deler av en validitetskjede vil det være preget av en høy kompleksitet, som kan inkludere alt fra læreplanen, skoleeieres forståelse av læreplanreform, lærerens operasjonaliseringer og elevenes handlinger i skolehverdagen. I denne studien vil jeg forsøke å se på sammenhengen mellom utvalgte aspekter ved validiteten i vurderingen i et bestemt kasus. De trinnene i validitetskjeden som vil undersøkes nærmere er *kompetansemål*, *vurderingssituasjon* og *vurdering*. En begrunnelse for valget av disse gis i kapittel 3.4.2.

I modellen av validitetskjeden til Crooks, Kane og Cohen (1996) er trusler for validitet knyttet opp til hvert element i kjeden, i motsetning til validitetskjeden presentert av Sandvik og Fjørtoft (2016) hvor det generelt nevnes hva som kan bidra til økt validitet. På bakgrunn av at jeg i denne studien vil analysere en prøve som en del av biologilærerens vurderingspraksis og ikke en standardisert prøve, er det mest hensiktsmessig å ta utgangspunkt i Sandvik og Fjørtofts (2016) modell. For å få en dypere innsikt i trusler for validitet som kan være aktuelle ved en summativ prøve som brukes formativt, vil jeg videre ta for meg ulike aspekter som Stobart (2012) har vektlagt ved formative vurderinger.

2.1.5 Validitet i klasseromskontekst

Stobart (2012) har tatt for seg trusler for validitet i formative vurderinger. Han definerer formative vurderinger som alle vurderinger der informasjon er samlet inn for å bidra til elevens læringsprosess. Med andre ord kan summative prøver som brukes formativt inngå. Først tar han opp tillit og motivasjon i sammenheng med om elevene tør å være delaktige i undervisning, og om det er et miljø hvor det er rom for å si feil. Disse momentene kan bidra til å øke eller redusere validiteten i den formative vurderingen. Dette kan ses i sammenheng med om elevene er aktive underveis, og om læreren får mulighet til å oppdage misoppfatninger og hvor elevene er i læringsprosessen. Får eksempelvis elevene mulighet til å utfordre forståelsen underveis i undervisningen?

Det neste poenget handler om å gjøre læringen eksplisitt. En sentral forutsetning i vurdering for læring er at elever og lærere må være tydelig på «hvor de skal gå videre». I praksis vil det si å være tydelig på hva som er formålet, og hva en god prestasjon er. Dette på bakgrunn av Sadlers (1989) oppfatning av at det er enklere å lære når vi vet hva vi lærer og kan evaluere vår prestasjon, men da må vi vite hva forventet standard er. Om dette ikke er tydelig, kan det være en trussel for validiteten i vurderingen. Samtidig påpeker Stobart (2012) at dette er en stram line å balansere på. Hvis læringsmålene blir spesifisert i for stor grad, kan det føre til fragmentert overflatelæring. Dette kan også bidra til å redusere validiteten. Torrance (2007) problematiserer dette. Han poengterer at om elevene vet nøyaktig hva som skal til for å få en bestemt karakter, og får veldig mye hjelp til å oppnå denne karakteren, så vil eleven gjøre suksess, men suksess for hva? Han mener dette kan bidra til å redusere kvaliteten og validiteten av det elevene oppnår.

En trussel for validiteten vil være om elevene ikke forstår hva de skal lære eller hvordan de kan nå læringsmålene. I tillegg, om elevene ikke har forståelse for formålet med undervisningen, vil det også kunne være vanskeligere for elevene å dra nytte av tilbakemeldingene som gis. Han understeker også i denne sammenheng viktigheten av å bryte ned kompetansemål til læringsmål eller eksempler og modeller for at de skal bli forståelige for elevene. Med andre ord kan uforståelige kompetansemål også utgjøre en trussel for validiteten.

Avslutningsvis omtaler Stobart (2012) tilbakemeldinger eller feedback. Hattie og Timperley (2007) har undersøkt kompleksiteten ved feedback, og hva som har innvirkning på om den har en positiv effekt. Blant annet vil nivå og kvalitet på tilbakemeldinger ha betydning, kompleksiteten på oppgavene og motivasjon hos den lærende. Stobart (2012) poengterer at to ulike individer som får samme tilbakemeldinger vil kunne ha ulik effekt, men at et viktig poeng med tilbakemelding er at den er myntet på oppgave fremfor person. Han trekker også frem at en karakter ikke sier noe om hva som bør gjøres videre i læringsprosessen, og at det fort kan være at eleven relaterer til å tenke at «jeg er denne karakteren» eller sammenligner seg med de andre i klassen. I tillegg vil elevenes bruk av tilbakemeldingene være av betydning. Hvis elevene ikke tar i bruk tilbakemeldingene i den videre læringsprosessen, vil det kunne være en trussel for validiteten i vurderingen.

Oppsummert vil spørsmålet være om vurderingen som gjennomføres bidrar til elevens læringsprosess. Denne gjennomgangen av trusler for validitet i formativ vurdering vil jeg bruke videre i analysen og diskusjonen knyttet opp mot validitetskjeden presentert av Sandvik og Fjørtoft (2016).

Bevissthet om validitetskjeden kan styrke gyldigheten i lærerens arbeid, og ifølge Black et al. (2010) kan validitetskjeden være et nyttig verktøy for å få innsikt i vurderingen i en praktisk kontekst. I denne studien vil det tas utgangspunkt i vurdering i biologifaget. Fagets egenart vil påvirke hva som regnes som gyldige bevis for læring (Sandvik & Fjørtoft, 2016), og derfor vil det være interessant å se nærmere på hva kompetanse i biologifaget innebærer og hvordan dette kan vurderes.

2.2 Vurdering av elevenes kompetanse i biologi

Vurdering og elevenes forventninger kan drive læring, og elever vil kunne selektivt studere og lære innhold og ferdigheter de tenker er essensielle for å gjøre det bra på prøven (Momsen et al., 2013). Ved å vurdere hva prøver i biologi tester, kan man få en innsikt i hva elevene vektlegger i læringsprosessen. Her vil jeg gå inn på hva som kjennetegner faget biologi i den norske skolen, kompetanse i faget og ulike kognitive ferdigheter. I tillegg vil jeg se på hva forskning sier om hvor sentral læreboken og eksamen er i biologiundervisningen i Norge.

2.2.1 Hva kjennetegner biologi som skolefag?

Biologifaget er komplekst. I sin enkleste definisjonsform kan en si at biologi er «læren om liv eller levende organismer» (Strømme, 2015), men hva kjennetegner egentlig biologi som fag? I boken «The Logic of Education» presenterer Hirst og Peters (1970) fire trekk som definerer et fag, nemlig begreper, syntaks, regler og karakteristiske ferdigheter, metoder, prosesser, redskaper og teknikker. Hvis vi bruker hovedområdet *genetikk* i læreplanen i biologi 2 som utgangspunkt, kan vi gjenkjenne disse trekkene? Innenfor dette området har vi egne *begreper* som for eksempel transkripsjon, translasjon, kodoner og arvestoff. Når det gjelder *syntaksen* er det regler for hva som regnes som legitime utsagn (Strømme, 2015). Eksempelvis er «alle barn der begge foreldrene har blå øyne vil få blå øyne» et legitimt utsagn i motsetning til «alle har en udødelig sjel». Observasjon og eksperimenter kan betraktes som *regler* for hvordan man avgjør om et utsagn er sant eller galt. Det siste punktet som omfatter eksempelvis *karakteristiske metoder*, kan innen genetikken være gentesting. Det kan diskuteres videre hva biologifaget egentlig innebærer, men i denne studien vil det være interessant å se videre på hva kompetanse i biologifaget er.

2.2.2 Kompetanse i biologi

I Kunnskapsløftet (LK06) defineres kompetanse som evnen til å løse oppgaver og mestre komplekse utfordringer (Utdanningsdirektoratet, 2016). I denne definisjonen ligger det implisitt en forståelse av at kunnskapene og ferdighetene ikke kun er et mål i seg selv, men at elevene må bruke kompetansen de tilegner seg til å løse oppgaver. Læreplanen i biologi er førende for hvilke kompetanser biologielevene skal oppnå, da kan det være interessant å stille seg spørsmål om hva er egentlig kompetanse i biologi? I denne studien er det hovedområdet

genetikk i læreplanen som det ses nærmere på, og derfor vil jeg analysere kompetansemålene innenfor dette området i kapittel 4.1.

2.2.3 Hvordan kan elevenes kompetanse i biologi vurderes?

Biologi kan betraktes som et mangfoldig fag som rommer aktiviteter som teoretiske studier, laboratorieøvelser og feltarbeid (Eggen & Vidnes, 2015). Biologifaget er generelt preget av mange begreper, og elevene må ha oversikt over mange ulike prosesser. Dette utgangspunktet har dermed betydning når elevenes prestasjoner skal vurderes, og det vil si at mange ulike vurderingssituasjoner kan og bør brukes. Biologi 2 regnes gjerne som et teoritungt fag hvor skriftlige prøver er vanlig, og elevene kan bli trukket opp i skriftlig eksamen. I denne studien vil oppgaver gitt på en skriftlig prøve analyseres. Prøven består av ulike typer oppgaver, derfor vil det være hensiktsmessig å beskrive disse typene.

2.2.4 Skriveoppgaver og flervalgsoppgaver

Oppgaver som gis på skriftlige prøver skal teste den kompetansen som prøven setter seg fore å teste (Angell et al., 2016). Oppgavene vil påvirke elevene til å rette oppmerksomheten mot bestemte aspekter av innhold, og dermed spesifisere måter å prosessere informasjon (Doyle, 1983).

Oppgavene på prøven i biologi 2 i denne studien deles inn i *flervalgsoppgaver* og *skriveoppgaver*. Skriveoppgaver vil her defineres som oppgaver der elevene må skrive tekst i besvarelsen. Flervalgsoppgaver vil være oppgaver der elevene skal angi alternativet som de anser som riktig svar på spørsmålet. Flervalgsoppgaver består av en stamme, også kalt en introduksjonssetning, ett riktig svaralternativ, og to eller flere gale svaralternativer som kalles distraksjoner (Brookheart, 2014). Flervalgsformatet har vært fremmed i norsk skole, men har etter hvert blitt innført som følge av at den har vært hyppig brukt på internasjonale og nasjonale prøver (Angell et al., 2016). Mange har stilt seg skeptisk til flervalgsoppgaver, og begrunnet det med at slike oppgaver kun tester faktakunnskaper og lavere ordens kognitive ferdigheter. Flere poengter derimot at dette oppgaveformatet utmerket kan utfordre elever på høyere nivå (Angell et al., 2016; Brookheart, 2014).

2.2.5 Bruk av lærebok i undervisningen

En studie av Bachmann (2005, s. 318) indikerte at over 80 % av lærere i den norske skolen planla undervisningen ut i fra læreboken. Andelen var størst i de naturvitenskapelige fagene. Den samme studien viste også at elever ofte jobbet med læreboktekst med tilhørende oppgaver i undervisningen. De samme tendensene viser rapporten gjennomført av Hodgson, Rønning og Tomlinson (2012). Hva kan så bruken av lærebok ha å si for validiteten i vurderingen i biologi? Et poeng kan være om læreboken brukes som utgangspunkt for hva som regnes som gyldig vurdering, og ikke kompetansemålene i læreplanen. En konsekvens kan være at kompetansemål som ikke prioriteres i læreboken, heller ikke prioriteres i vurderingssituasjoner. Hva om elevene kan oppnå full score på prøven ved å gjengi faktakunnskaper fra læreboken? Dette setter søkelyset mot validiteten i vurderingen og bruk av lærebok i undervisningen.

2.2.6 Er eksamen styrende?

Ifølge Raaheim og Hauge (2007) er eksamen styrende for hvordan lærere legger opp undervisningen. I en rapport om naturfagene i den norske skolen skrevet på oppdrag fra Utdanningsdirektoratet (2015), oppga over 40 % av lærere i biologi at eksamen var styrende for både innholdet, arbeidsformene og vurderingsformene i faget. En konsekvens av dette kan være at eksamen brukes som en tolkning av læreplanen. Dette kan igjen føre til at de kompetansemålene som sjeldent blir testet på eksamen, ikke får plass i undervisningen og dermed ikke blir vurdert (Utdanningsdirektoratet, 2015). Hvis elevene får definert nøyaktig hva de skal kunne, kan det hindre dem i stille de store spørsmålene og søke kunnskap utenfor det området som er definert (Fjørtoft, 2016). Med andre ord kan validiteten i vurderingen påvirkes.

2.3 Kognitive ferdigheter

Beskrivelse av kompetanse på ulike nivåer har tradisjonelt vært knyttet til spørsmål om taksonomi i den norske skolen (Thronsen, Hofpenbeck, Lie & Dale, 2009). I den hensikt å dele kompetansemål og prøveoppgaver inn etter hvilke kognitive ferdigheter de krever vil jeg gi en nærmere beskrivelse av henholdsvis Blooms taksonomi og Webbs nivåinndeling.

2.3.1 Blooms taksonomi

Et kjent klassifiseringsverktøy for pedagogiske mål er Blooms taksonomi (Anderson et al., 2001; Bloom, 1956). Han delte kunnskapsfeltet inn i tre hovedområder som omhandlet det kognitive, det affektive og det psykomotoriske. Det kognitive området tar for seg kunnskaper. Det affektive området handler om holdninger og verdier, mens det psykomotoriske området tar for seg fysiske ferdigheter. Videre vil inndelingen av det *kognitive området* presenteres nærmere, da det danner utgangspunktet for analysen av prøveoppgavene.

Det kognitive området klassifiserer ulike krav til kompetanse etter økende grad av kompleksitet når det gjelder oppgaveløsning (Bloom, 1956). Nivåene omtales som hierarkiske, det vil si at et høyere nivå forutsetter nivåene under. I Tabell 1 presenteres de ulike nivåene der *evaluering* med høyest kompleksitet er plassert øverst, deretter *syntese*, *analyse*, *anvendelse*, *forståelse* og nederst er *kunnskap* med lavest kompleksitet.

Tabell 1. Forenklet utgave av Blooms taksonomi der kunnskap beskriver lavest kompleksitet, og evaluering høyest kompleksitet (tilpasset fra Eggen og Vidnes, 2015, s. 241).

Nivå i Blooms taksonomi	Kjennetegnes ved at elevene kan
Vurdering	Bedømme ut fra ulike kriterier, drøfte, forsvare, avgjøre, forkaste, godta, sammenligne
Syntese	Fortolke eller gå fra deler til helhet, oppsummere, presisere, organisere, trekke slutninger
Analyse	Se i system, gå fra helhet til deler, velge ut, klassifisere
Anvendelse	Gjøre rede for og bruke kunnskap i nye sammenhenger
Forståelse	Forklare med egne ord, enkle tolkninger
Reproduksjon	Gjengi, gjenkjenne faktakunnskaper

Denne taksonomiske modellen har fått kritikk fra flere hold fordi den ikke bygger på empiri, og at den baserer seg på et forenklet syn på tenking og hvordan det henger sammen med

læring (Marzano & Kendall, 2007). Samtidig peker Eggen og Vidnes (2015) på den lange tradisjonen med bruk av Blooms taksonomi i skoleverket i Norge. De bemerker også at kompetansemålene i læreplanen og at enkelte eksamensoppgaver kan graderes etter denne taksonomien. Med andre ord kan dette rettferdiggjøre hvorfor Blooms taksonomi brukes som utgangspunkt for et analyseverktøy i denne studien. Samtidig belyser det hvorfor jeg har sett nødvendigheten av å kombinere Blooms taksonomi med Webbs nivåinndeling som analyseverktøy.

2.3.2 Webbs nivåinndeling

Det har blitt utviklet flere klassifiseringssystemer, og en av dem er Webbs Depth of Knowledge (Webb, 1997). Modellen består av fire nivåer der kompetansemål eller oppgaver kan plasseres ut i fra dybde og kompleksitet. Dybde vil i denne sammenheng si hvilke kognitive prosesser kompetansemålene eller oppgavene legger opp til hos eleven (Webb, 1997). Kjennetegn på de ulike nivåene er oppgitt i Tabell 2. Nivå 1 kalt *gjenkjennelse og reproduksjon* krever at eleven bruker enkle ferdigheter som å gjengi fakta, som for eksempel gjengi definisjoner eller bruke enkle prosedyrer. Nivå 2 som er *grunnleggende bruk av begreper* krever at eleven bruker kunnskapen utover det som er beskrevet i første nivå. Eksempelvis at eleven må ta beslutninger om hvordan et problem skal løses, ifølge Webb (1997) innebærer nivå 2 at eleven viser forståelse. Nivå 3 som i Tabell 2 er kalt *strategisk tenkning* tar for seg argumentasjon og planlegging, eksempelvis at eleven må koble sammen ulike ideer eller forklarer fenomener ved å bruke kunnskapen i en gitt sammenheng. Et annet eksempel er at eleven bruker observasjoner til å utvikle en konklusjon. Det siste nivået, *utstrakt tenkning*, uttrykker kompleks tenkning og argumentasjon. Eksempelvis å se sammenhenger mellom ideer på tvers av ulike tema, og dermed måtte ta hensyn til flere variabler.

Tabell 2. Webbs nivåinndeling etter hva som kreves av elevene (tilpasset fra Webb, 1997).

Webbs inndeling	Kjennetegnes ved at elevene kan
Utstrakt tenkning	Kompleks argumentasjon, utvikling og tekning. Relatere ideer på innen og på tvers av tema. Ta hensyn til flere variabler.
Strategisk tenkning	Kan innebære argumentasjon og planlegging. Eleven kan gjerne oppfordres til å forklare sin tekning eller koble sammen ulike ideer. Oppgavene krever at elevene kan konkludere fra observasjoner, forklare fenomener ut i fra deres kunnskap og bruker kunnskapen til å løse ulike utfordringer.
Grunnleggende bruk av ferdigheter/begreper	Krever forståelse og bruk av kunnskap utover det å gjenkjenne og reprodusere. Det kan innebære at eleven må ta noen beslutninger om hvordan et problem skal løses.
Gjenkjennelse og reproduksjon	Krever reproduksjon eller gjenkjennelse av informasjon. Bruker kun enkle ferdigheter. Eksempelvis definisjoner, formelbruk og enkle prosedyrer.

Videre i kapittel 3.4.1 presenteres et analyseverktøy som baserer seg på Blooms taksonomi og Webbs nivåinndeling. Hensikten med å kombinere disse inndelingene er at Webbs nivåinndeling i større grad beskriver konteksten sammenlignet med Blooms taksonomi, og at de i så tilfelle kan utfylle hverandre som analyseverktøy.

I kapittel 2 har jeg gått gjennom vurdering, validitetskjeder og trusler for validitet i vurdering. I neste delkapittel så jeg nærmere på vurdering av elevenes kompetanse i biologi, samt forskning på bruk av lærebok og om eksamen er styrende for biologiundervisningen ble tatt opp med tanke på trusler for validitet. Avslutningsvis beskrev jeg hvordan man kan bruke Blooms taksonomi og Webbs nivåinndeling for å undersøke hvilke kognitive ferdigheter som kreves på kompetansemål og oppgaver. Videre i kapittel 3 vil jeg beskrive aspekter ved den metodiske tilnærmingen i studien.

3 Metodisk tilnærming

I dette kapitlet vil jeg beskrive og begrunne metodevalg som er gjort for å kunne besvare forskningsspørsmålene. Jeg har tatt utgangspunkt i en kvalitativ innholdsanalyse av kompetansemål og en skriftlig prøve i hovedområdet *genetikk* i biologi 2. I tillegg er det gjennomført et fokusgruppeintervju med elever i biologi 2 for å få en innsikt i deres opplevelser og erfaringer med vurderingspraksisen til biologilæreren. Jeg intervjuet også biologilæreren for å få bakgrunnsinformasjon om vurderingspraksisen hans, og dette brukes som sekundær empiri. Først vil jeg presentere forskningsdesignet, deretter aspekter ved datainnsamling og analysemetodene som er delt inn etter det som omhandler innholdsanalysen og fokusgruppeintervjuet. Avslutningsvis vil metodekvalitet og etiske betraktninger legges frem.

3.1 Kontekst

Forskningen ble gjort i februar og mars 2018 i en biologi 2-klasse. Utvalget besto av fem elever og en lærer, i tillegg til dokument i form av en prøve i genetikk og kompetansemål. Skolen har deltatt i et prosjekt kalt «SKUV – skolebasert kompetanseutvikling i vurdering», og har dermed hatt fokus på en styrket vurderingskompetanse og god vurderingspraksis. Prosjektperioden er over fire år, og hver skole jobber med et tema over tre semestre. I denne perioden skal de som deltar jobbe kollektivt i hver enkelt skole for å forbedre og utvikle vurderingskompetansen, og skoleforskere utenifra bidrar i denne prosessen. Biologilæreren har dermed hatt fokus på vurdering i forkant av forskningen jeg har gjort. Han har tatt en master i biologi, bygd på med pedagogikk og årsstudium i matematikk. Han har jobbet over ti år i den videregående skolen. I intervjuet kom det frem at undervisningen i stor grad følger læreboken, og at det er ganske vanlig med forelesninger. Lærere fortalte også at ved utforming av prøver reflekterer han over å stille spørsmål som dekker de områdene de har vært gjennom i undervisningen, og at det er viktig at elevene får trening i eksamensoppgaver da de ofte er annerledes enn «hva er» og «forklar det og det».

Prøven som analyseres regnes som en «typisk» prøve av elevene, der de får et sett flervalgsoppgaver og skriveoppgaver. Maks antall poeng var ikke oppgitt på oppgavearket, men elevene fortalte at de tidligere har blitt gitt et halvt poeng når de har angitt riktig svaralternativ på flervalgsoppgavene. Skriveoppgavene gir vanligvis høyere uttelling.

3.1.1 Metodisk rammeverk

Studiens rammer bygger på sosialkonstruktivistiske teorier hvor det tas utgangspunkt i at virkeligheten konstrueres og rekonstrueres gjennom handlinger og interaksjoner mellom mennesker (Berger & Luckmann, 1991). Virkeligheten kan bli endret og forbedret av mennesker, og dermed har jeg gjennom denne studien forsøkt å få innsikt i forskningsdeltagernes oppfatninger av deres virkelighet. Ved hjelp av deres refleksjoner og mine tolkninger har jeg ønsket å utvide forståelsen innenfor forskningsfeltet.

3.2 Forskningsdesign

Studien tar sikte på å gå i dybden av undersøkelsene, noe som kjennetegner forskning innenfor den kvalitative forskningstradisjonen (Robson & McCartan, 2016). Dette er i motsetning til kvantitative metoder som fokuserer på generelle sannheter basert på statistiske analyser (Christofferen & Johannessen, 2012). Formålet med studien har vært å forstå fremfor å forklare, og den kan dermed plasseres innen det fortolkende forskningsparadigmet (Husén, 1988).

Forskningstilnærmingen kan karakteriseres som en kausstudie. Slike studier kjennetegnes av at det forskes på få personer eller dokumenter, den er begrenset i tid og det benyttes flere datakilder (Yin, 2014). I denne kausstudien forskes det på vurderingspraksisen i en biologi 2-klasse, og datakildene består av fokusgruppeintervju med fem elever, en kapittelprøve i genetikk og kompetansemål. I tillegg brukes intervjuet med biologilæreren som sekundær empiri. Begrunnelsen for dette er at jeg ønsket at elevenes refleksjoner skulle stå i sentrum.

3.3 Datainnsamling

Dokumenter som datakilde

Innholdsanalyse er en av analysemetodene som benyttes i denne studien. Metoden kjennetegnes blant annet av at datakilden er skriftlige dokumenter. Jeg har tatt utgangspunkt i *directed content analysis* (Hsieh og Shannon, 2005). I denne analysemetoden analyseres dokumenter etter at jeg som forsker har satt meg inn i teoretisk rammeverk og tidligere forskning. Deretter utvikles koder som brukes som verktøy for å analysere dokumentene.

Utvalg til innholdsanalyse

Dokumentene som har blitt analysert består av læreplan i biologi 2 og en prøve gitt i en biologi 2-klasse våren 2018. Alle delene som har blitt analysert er i hovedområdet *genetikk*. Kompetansemålene for genetikk ble analysert, da de er beskrivende for hva elevene skal kunne etter endt skolegang.

Fokusgruppeintervju som datakilde

Fokusgruppeintervju ble gjennomført for å få frem elevenes refleksjoner om vurderingspraksisen i biologiklassen. Denne kvalitative forskningsmetoden ble brukt fordi gruppeinteraksjonene kan by på rike svar ettersom deltakerne stimuleres av tankene og kommentarene til de andre i gruppen (Robson & McCartan, 2016).

Det ble utarbeidet en intervjuguide på forhånd som kan ses i Vedlegg B, med fastsatte tema som et utgangspunkt for å belyse vurderingpraksisen. Gjennomføringen av intervjuet bar preg av det Christoffersen og Johannessen (2012) kaller semistrukturerte intervjuer, og ga dermed mulighet til å avvike fra intervjuguiden når interessante aspekter dukket opp underveis. En viss struktur ble holdt, men samtidig var det viktig å utdype temaer som elevene valgte å trekke frem. Min rolle som moderator ble å «lese» diskusjonen og sørge for at alle kom til ordet, nettopp slik som Malterud (2012) anbefaler.

Utvalg til intervju

Utvalget til intervjuene i denne studien bestod av fem elever og en lærer fra en biologi 2-klasse. Elevene har fått de fiktive navnene: Trine, Nora, Beate, Johanne og Jonas. Intervjuet med læreren ble gjennomført for å samle inn bakgrunnsinformasjon til studien, og er dermed ikke analysert i motsetning til fokusgruppeintervjuet med elevene. Jeg hadde ingen relasjon til læreren eller elevene før kontakt ble opprettet under oppstart av studien. Utvalget av elevene kan kalles et tilgjengelighetsutvalg da det var disse fem som selv meldte seg til å delta. Læreren stilte seg til disposisjon som forskningsdeltaker, både ved å la seg intervju og gi tilgang på de skriftlige prøvene han hadde utformet og gjennomført.

3.4 Analyseverktøy og analysetrinn

Her vil jeg gi en beskrivelse av analyseverktøyet som ble brukt i innholdsanalysen av kompetansemålene og prøven. I tillegg vil jeg forklare nærmere analysetrinnene i innholdsanalysen og analysen av transkripsjonen av fokusgruppeintervjuet.

3.4.1 Analyseverktøy og analysetrinn i innholdsanalysen

For å besvare forskningsspørsmålene knyttet til kompetansemålene og prøven var det nyttig å ha fastsatt kodene på forhånd. Disse kodene ble utarbeidet med utgangspunkt i Blooms taksonomi og Webbs nivåinndeling. I Tabell 3 presenteres inndelingen av de ulike kategoriene.

Tabell 3. Tabellen viser inndeling av kategoriene i innholdsanalysen basert på Blooms taksonomi og Webbs nivåinndeling. B1-B6 indikerer Blooms inndeling fra henholdsvis reproduksjon som minst komplekst, til vurdering som mest komplekst. W1-W4 viser til nivå 1 til nivå 4 i Webbs nivåinndeling.

Klasse	Beskrivelse	Eksempler på verb
LAV (B1, B2, W1, W2)	<ul style="list-style-type: none">• Faktakunnskaper• Gjengivelse	Gjengi, beskrive, måle, beregne, oppsummere, definere, liste opp, bruke, forklare, sammenligne, gjøre rede for, fortell
MIDDELS (B3, B4, W2, W3)	<ul style="list-style-type: none">• Forståelse til å anvende kunnskap• Prosesser som omhandler flere steg• Se sammenhenger	Utføre, organisere, tolke, analysere
HØY (B5, B6, W4)	<ul style="list-style-type: none">• Bruke kunnskap i komplekse prosesser• Trekke konklusjoner• Foreslå fremgangsmåter• Se sammenhenger på tvers av tema	Bedømme, drøfte, begrunne, planlegge, foreslå, vurdere

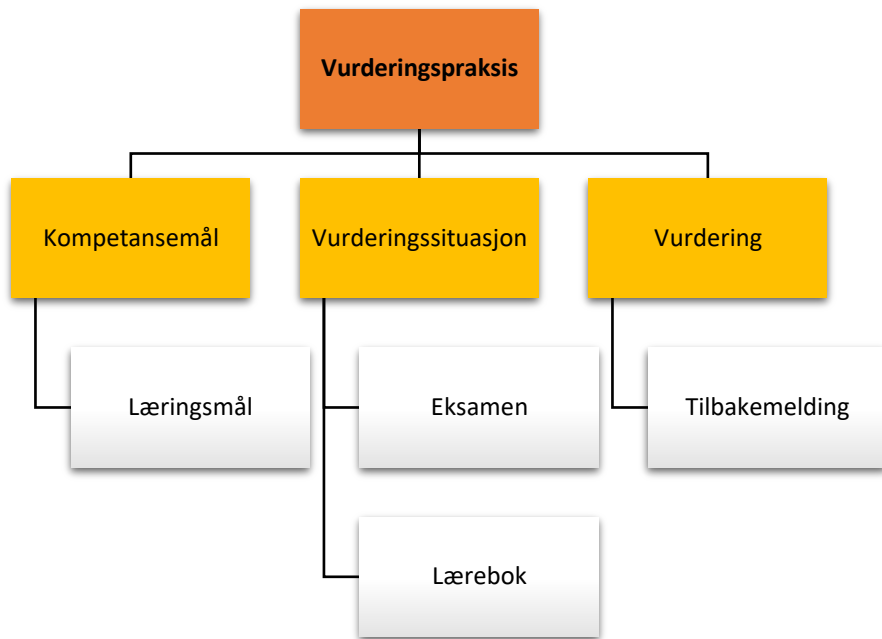
Innholdsanalysen ble delt i to, det vil si at kompetansemål og prøveoppgavene ble analysert hver for seg for å se hvilke kognitive ferdigheter de krevde av elevene. Ved hjelp av verbene som ble brukt i kompetansemålene og konteksten, ble de plassert innenfor de ulike kategoriene

presentert i Tabell 3. Prøveoppgaven ble delt opp i del 1 som var flervalgsoppgavene og del 2 som var skriveoppgavene. På flervalgsoppgavene var det også viktig å vurdere disktraktorene som ble brukt.

3.4.2 Analysemetode og analysetrinn i fokusgruppeintervjuet

Analysemetoden som ble brukt for å fremstille og analysere resultatene fra fokusgruppeintervjuet har tatt utgangspunkt i det Robson og McCartan (2016) kaller tematisk koding. I starten av analyseprosessen transkriberte jeg lydopptaket til tekst. Transkripsjonen ble lest gjennom flere ganger, og under disse prosessene ble tanker og ideer av interesse i forbindelse med problemstillingen notert. Etter å ha kodet elevenes ytringer fra fokusgruppeintervjuet plasserte jeg kodene i følgende kategorier: *kompetansemål*, *vurderingssituasjon* og *vurdering*, som alle er deler av validitetskjeden.

Elevene hadde ikke kjennskap til læringsmål, vurderingskriterier eller kjennetegn på måloppnåelse i den aktuelle vurderingssituasjon. I tillegg var ikke arbeidsoppgaver mye omtalt under intervjuene, derfor var det naturlig å sette søkelys på kompetansemål, vurderingssituasjon og vurdering. Jeg opplevde at disse kategoriene rommet mange aspekter som elevene knyttet til temaet. Sammenheng mellom kategoriene og tilhørende koder illustreres i flytskjemaet i Figur 3.



Figur 3. Flytskjemaet illustrerer hovedkategoriene *kompetansemål*, *vurderingssituasjon* og *vurdering* (gul) med tilhørende underkategorier (hvit). Disse er blitt utarbeidet fra transkripsjonen av fokusgruppeintervjuet med elevene i biologi 2.

3.5 Metodekvalitet

Begrep som validitet og reliabilitet har tradisjonelt blitt brukt til å si noe om gyldigheten og påliteligheten til en studie, og har med tiden blitt overført fra kvantitativ til kvalitativ forskning (Robson & McCartan, 2016). Guba (1981) introduserte begrepene *kredibilitet* (credibility), *overførbarhet* (transferability), *avhengighet* (dependability) og *bekreftbarhet* (confirmability) i den hensikt å si noe om troverdigheten til kvalitative studier. Denne studien tilhører det fortolkende forskningsparadigmet, og dermed er det en fare for at subjektivitet kan påvirke resultatene. Dette er et argument for at gi redegjørelser av hva som har blitt gjort for å styrke metodekvaliteten.

3.5.1 Kredibilitet

Kredibilitet omhandler hvordan det kan etableres tillit om at funnene fra studien er holdbare og sanne (Guba, 1981). Metodekvaliteten kan styrkes ved å gjøre «member checks» (Robson & McCartan, 2016), et eksempel på dette er forskningsdeltagerne får tilgang til transkripsjonene som er skrevet med utgangspunkt i lydopptakene av intervjuene som er gjort. Alle

forskningsdeltagerne har fått tilsendt transkripsjonene av intervjuene med mulighet til å gi tilbakemelding. Robson og McCartan (2016) foreslår også «peer debriefing» med andre forskere. Dette har blitt gjennomført ved at tanker og deler av tekst har blitt delt med medstudenter, noe som har ført til verdifulle diskusjoner. I tillegg kan triangulering av datakilder bidra til å øke kredibiliteten, og i denne studien er det brukt flere datakilder.

3.5.2 Overførbarhet

Når det gjelder kasusstudier innenfor det naturalistiske forskningsparadigmet trekker en gjerne frem overførbarhet fremfor generalisering, dette på bakgrunn av den sosiale konteksten som har innvirkning på kasuset (Guba, 1981). Overførbarhet vil i dette tilfellet si noe om funnene fra denne studien kan anvendes i en ny og annen kontekst. Et bidrag kan være «tykke beskrivelser» av forskningens kontekst, slik at den som leser oppgaven kan avgjøre om det gir grunnlag for å generalisere. I kapittel 3.1 har jeg forsøkt å gi en tykk beskrivelse av konteksten til studien som kan brukes som bakgrunnsinformasjon av leseren.

3.5.3 Avhengighet

Når det gjelder avhengighet kan det stilles spørsmål om andre forskere ville fått de samme resultatene, og det er her snakk om konsistens. I denne studien har det blant annet blitt brukt et analyseverktøy basert på Blooms taksonomi og Webbs nivåinndeling av kategorier. Når kompetansemål og prøveoppgaver vurderes etter hvilke kategorier de tilhører, vil det være en tolkning gjort av meg. Spesielt når det gjelder tolkninger av relativt åpne kompetansemål, vil ulike forskere og lærere kunne komme frem til ulike kategorier. Med andre ord er det ikke en garanti for at andre forskere ville fått nøyaktig samme resultat. Samtidig er det forsøkt å gjøre analyseverktøyet så grundig som mulig ved å bruke både verb og kontekst som utgangspunkt. Det er i tillegg forsøkt å gi en god beskrivelse av datainnsamling og dataanalyse, slik at andre forskere har mulighet til å se og vurdere mine refleksjoner rundt prosedyrer og avgjørelser.

3.5.4 Bekreftbarhet

Bekreftbarheten tar for seg objektiviteten av studiet, med andre ord om funnene ville blitt repetert om studien ble gjentatt i tilsvarende kontekst med tilsvarende deltakere og/eller dokumenter. Et eksempel som kan bidra til objektivitet er at analytiske påstander bygges opp

av datakilder og teori, noe som er gjort i kapittel 5. En annen faktor som kan ha positiv innvirkning på dette området er transparens og åpenhet, noe som er forsøkt å oppnå i beskrivelsen av hvordan resultatene har fremkommet. I tillegg har jeg redegjort for studiens rammer med tanke på sosialkonstruktivisme.

3.6 Ethiske betraktninger

Et viktig aspekt ved kvalitativ forskning er etikk (Robson & McCartan, 2016), og det er noe jeg har forsøkt å være bevisst på under hele prosessen. Da deler av datainnsamlingen skulle behandle personopplysninger var studien meldepliktig, og ble derfor meldt inn til Norsk senter for forskningsdata (NSD). Godkjennelsen av studien kan ses i Vedlegg D. Det ble foretatt små endringer etter første innsending, men etter å ha kontaktet NSD fikk jeg tilbakemelding om at endringene ikke var tilstrekkelig stor for at ny søknad var nødvendig.

NSD stiller krav om at utvalget som intervjues må informeres om prosjektet og gi samtykke til deltakelse. Informasjonsskrivet til elevene og læreren ble utformet ved hjelp av de forskningsetiske retningslinjene fastsatt av Den nasjonale forskningskomite for samfunnsvitenskap og humaniora (NESH, 2016). Biologilæreren fikk tilsendt samtykkeskrivet med informasjon slik at han kunne høre med biologiklassen hvilke elever som kunne være interessert i å delta i studien. Samtykkeskrivet kan ses i Vedlegg E. Før gjennomføringen av studien ble elevene igjen påmint å lese gjennom samtykkeskrivet, i tillegg til en muntlig orientering om av det var helt frivillig å delta, samt at de kunne trekke seg fra undersøkelsen når som helst uten å oppgi årsak. Det var også viktig å poengtere at navnene deres ville bli anonymisert både underveis i arbeidet med masteroppgaven og ved endelig publisering. Alle elevene var over 15 år, og kunne dermed selv samtykke til at jeg kunne innhente og bruke deres personopplysninger (NESH, 2016). Det var fem elever og én lærer som skrev under samtykkeskjemaet, og dermed ble gjort lydopptak av.

Under hele prosessen har jeg vært oppmerksom på å prøve og gi en så nøytral gjengivelse som mulig av de datamaterialer som er samlet inn. Alle forskningsdeltakere har fått fiktive navn, og verken skole eller lærer har blitt omtalt med navn eller annen informasjon som kan være identifiserende. Samtidig er det viktig å ta min rolle som forsker til betraktning. Store deler av analysearbeidet har gått ut på å plassere tekst i form av kompetansemål, prøveoppgaver og

koder fra intervju under ulike kategorier ut ifra min tolkning. Tidligere nevnte jeg at en fordel med innholdsanalyse er at forskeren ikke har innvirkning på dokumentet det forskes på, men samtidig kan min tolkning av hvor for eksempel oppgaver bør plasseres med tanke på de ulike kategoriene tas i betraktning. Med andre ord er det ikke sikkert at andre forskere hadde kommet frem til nøyaktig samme resultat. Resultatet kan bære preg av min bakgrunn som lektorstudent i realfag, samtidig har jeg prøvd å være mest mulig objektiv ved å bruke det teoretiske rammeverket som utgangspunkt.

Et annet aspekt å ta hensyn til er at det kan være fare for at biologilæreren kan føle seg ekstra utsatt for kritikk siden jeg kun har analysert en biologilærers vurderingspraksis. Det har dermed vært viktig for meg å presisere at jeg på ingen måte vil kritisere vurderingspraksisen jeg har forsket på, men at den kan være et viktig bidrag til å få mer kunnskap om en av mange forskjellige vurderingspraksiser i biologi 2.

4 Analyse

I dette kapittelet vil jeg presentere resultatene fra analysen, og bruke eksempler fra empiri for å underbygge de analytiske påstandene jeg legger frem. Først vil datamaterialet fra innholdsanalysene av kompetansemålene og prøven i genetik beskrevet for å svare på det første forskningsspørsmålet. Deretter vil biologielevenes refleksjoner vektlegges med utgangspunkt i hovedkategoriene *kompetansemål*, *vurderingssituasjon* og *vurdering*, presentert i Figur 3 for å besvare det andre forskningsspørsmålet.

4.1 Kompetansemålene indikerer bruk av lavere og middels kognitive ferdigheter

Med bakgrunn i at læreplanen er dokumentet som gir retningslinjer for hva som skal undervises i biologi 2, var det interessant å analysere kompetansemålene gitt i hovedområdet *genetik* ved hjelp av analyseverktøyet basert på Blooms taksonomi og Webbs nivåinndeling. Årsaken til at nettopp disse kompetansemålene ble valgt som utgangspunkt for analysen, var at elevene jobbet med hovedområdet *genetik* i perioden forskningen pågikk.

Kompetansemålene i hovedområdet *genetik* biologi 2:

- 1) Forklare strukturen til DNA og hvordan DNA blir kopiert før cellene deler seg.
- 2) Gjøre greie for transkripsjon og translasjon av gen og forklare hvordan regulering av gen kan styre biologiske prosesser.
- 3) Sammenligne mitose og meiose med vekt på fordelingen av genmateriale i cellene som blir dannet.
- 4) Sette opp og teste hypoteser for kjønnsbunden og dihybrid arvegang med og uten kobling av gen
- 5) Forklare genetiske sykdommer ved å bruke kunnskaper om arv og mutasjoner, og gjøre greie for hvordan samspillet mellom arv, miljø og livsstil kan påvirke helsa hos mennesker. (Udir., 2013)

Kompetansemålene som er listet opp over viser at verb som «å forklare», «gjøre greie for» og «sammenligne» går igjen. Alle disse verbene går inn under kategorien *lav*, da de stort sett

inngår i å kunne gjengi faktakunnskaper. Hvis man ser nærmere på kompetansemålet «Gjøre greie for transkripsjon og translasjon av gen og forklare hvordan regulering av gen kan styre biologiske prosesser» er det blitt klassifisert i kategorien *lav*, da eleven blir bedt om å gjøre greie for og forklare (merket gult i eksemplet under). Samtidig er det veldig mange begrep eleven må ha forståelse for å kunne forklare prosessen, eksempelvis RNA, mRNA, tRNA, rRNA, RNA-polymerase, pre-mRNA, eksoner, introner, spleisosomer, kodoner, antikodon og ribosomer. Eleven blir ikke bedt om å bruke kunnskapen i større sammenhenger, og kompetansemålet ble derfor kategorisert som at det ikke trengs kunnskaper utover det lavere kognitive nivået.

- 2) Gjøre greie for transkripsjon og translasjon av gen og forklare hvordan regulering av gen kan styre biologiske prosesser.²

Ett av kompetansemålene som skiller seg ut fra de andre er «sette opp å teste hypoteser for kjønnsbunden og dihybrid arvegang med og uten kobling av gen». Dette har jeg plassert innenfor kategorien *middels*, da elevene skal bruke kunnskapen om kjønnsbunden og dihybrid arvegang i en større sammenheng. Eleven bruker kunnskapen for å kunne teste hypoteser som han/hun selv setter opp. Her er det også mange begreper eleven må ha kjennskap til for å kunne utføre slike oppgaver.

Det siste kompetansemålet som omhandler genetiske sykdommer, mutasjoner, arv, miljø og livsstil har jeg også plassert innenfor kategorien *middels*. Dette fordi eleven må bruke kunnskapen i en større sammenheng, og kunne koble ideer mellom ulike temaer i genetikken. Selv om verbene «å forklare» og «gjøre greie» for brukes her, mener jeg at konteksten indikerer bruk av kunnskaper utover lavere kognitive ferdigheter.

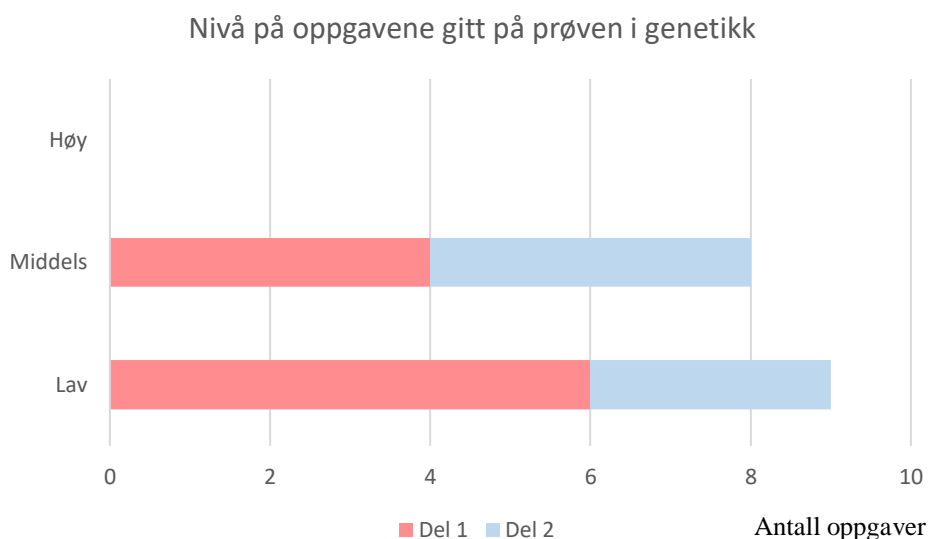
Samlet sett har de tre første kompetansemålene blitt plassert innenfor kategorien *lav*, mens de to siste ble plassert innenfor kategorien *middels*. Samtidig er det viktig å presisere at

² Verbene i kompetansemålet er merket med gult.

kompetansemålene kan oppfattes ulikt fra person til person med tanke på hvilke ferdigheter de krever, så dette kan ikke betraktes som eksakt vitenskap.

4.2 Prøveoppgavene indikerer bruk av lavere og middels kognitive ferdigheter

Når det gjelder hvilke kognitive nivå oppgavene på prøven krevde av elevene, er det gitt en oppsummering i Figur 4. Den første delen inneholder alle flervalgsoppgavene på prøven, mens del to tar for seg skriveoppgavene. Alle oppgavene på prøven er plassert innenfor kategoriene *lav* og *middels*, med andre ord er ingen av oppgavene plassert i kategorien *høy* som blant annet kjennetegnes av å bruke kunnskap i komplekse prosesser, og se sammenhenger på tvers av tema. Figuren indikerer at åtte oppgaver er plassert i kategorien *middels*, og ni oppgaver er plassert i kategorien *lav*.



Figur 4. Figuren viser hvilke kategorier de ulike oppgavene på prøven er plassert i. Flervalgsoppgavene er plassert i del 1, mens del 2 tar for seg skriveoppgavene.

I og med at dette er en kvalitativ studie vil jeg gi en utfyllende beskrivelse av et utvalg oppgaver som er plassert innen henholdsvis kategorien *lav* og *middels*. Oppgavetekstene vises som utklipp fra det opprinnelige prøvesettet som kan ses i Vedlegg A.

4.2.1 Eksempler på oppgaver som er plassert innen kategorien *lav*

Ni av i alt 17 oppgaver er plassert innen kategorien *lav*, og her vil jeg gå gjennom et utvalg av eksempler med tilhørende beskrivelser. I den første oppgaven er spørsmålet hva man kaller egenskapen å ha to identiske alleler i en organisme. Her må elevene velge mellom fire ulike alternativer som alle er ulike begreper som brukes i genetikken. Oppgaven har blitt plassert i kategorien *lav* fordi den krever at eleven bruker faktakunnskaper uten å anvende den i større sammenhenger.

1. En organisme med to identiske alleler for en egenskap blir kalt
 - homozygot
 - dominant
 - selvpollinerende
 - heterozygot

I den tredje oppgaven i prøvesettet skal elevene vite hva som representerer en testkrysning. Denne oppgaven er også plassert i kategorien *lav*, da den ikke krever kunnskap utover det å gjenkjenne.

3. Hvilken krysning er eksempel på en testkrysning?
 - TT × Tt
 - Aa × Aa
 - Dd × dd
 - Pp × PP

Den 9. oppgaven krever at elevene har kjennskap til replikasjon av DNA, og hva som er viktig for å bevare DNA-sekvensen. Heller ikke i denne oppgaven må elevene anvende kunnskapen i en større sammenheng. Hvis de har pugget det som står i læreboken vil de kunne angi korrekt svaralternativ.

9. Hva av dette er essensielt for å bevare DNA-sekvensen ved replikasjon?
 - Ett gen koder for ett polypeptid
 - Komplementær baseparing
 - Substrater er spesifikke for enzymer
 - Den genetiske koden er universell
 - Fotofosforylering

Det siste eksemplet er hentet fra del to på prøven som består av skriveoppgaver. Her skal elevene ha kjennskap til bruk av krysningsskjema, og hvordan de finner fenotype- og genotypefrekvens. Denne deloppgaven er også plassert i kategorien *lav*, da den legger opp til teknisk bruk av kunnskapen.

11. Pelsfarge hos rødvrev bestemmes av to gener. Begge genene må ha minst ett dominant allel for at den røde pelsen skal komme til uttrykk, hvis ikke får man sølvfarget pelsfarge.

a) Dersom du krysser to sølvrever med genotypene $ssTT$ x $SStt$ hva blir da utfallet? Vis kryssning, samt fenotype-/genotypefrekvens.

Samlet sett viser utvalget av oppgaver hva som kreves i kategorien kalt *lav*.

4.2.2 Eksempler på oppgaver som er plassert innen kategorien *middels*

Åtte av 17 oppgaver er plassert innenfor kategorien *middels*. Jeg vil her gå gjennom eksempler på oppgaver som er plassert innenfor denne kategorien. Den første oppgaven er en flervalgsoppgave der eleven skal avgi korrekt alternativ på spørsmål om fordelingen av avkommet ved kryssning mellom to blå petuniaer.

5. En kryssning mellom en purpurfarget petunia og en hvit petunia ga 100 avkom som alle var blå. Når to blå petuniaer krysses med hverandre blir fordelingen i avkommet:
- 75 % hvite, 25 % purpur
 - 75 % purpur, 25 % hvite
 - 100 % blå
 - 25 % purpur, 50 % blå, 25 % hvite

Her er det tatt utgangspunkt i at eleven får oppgitt informasjon, som han/hun må bruke videre i sine antagelser om hvilket utfall kryssningen vil gi. Med andre ord kunne oppgaven plasseres i kategorien *middels*. De to neste oppgavene som vises som utklipp merket med d) og e) er tatt fra en større oppgave der elevene får oppgitt at pelsfargen hos rødvrev bestemmes av to gener, som begge må ha minst ett dominant allel for at den røde pelsen skal komme til uttrykk. I oppgaveteksten står det videre at hvis dette ikke er tilfellet, vil man få en sølvfarget pelsfarge. Elevene blir så spurt om hvor stor sjans det ville vært å få sølvrever om genene er koplet, og en kryssning mellom to heterozygote med genotyper $SsTt$ x $SsTt$ ble krysset.

d) Hvordan ville det sett ut om genene var koplet og to heterozygote med genotyper $SsTt$ x $SsTt$ ble krysset?

I neste deloppgave følges så forrige spørsmål opp med å anta koplete gener med en overkrysning hos et av individene med den samme genotypen, og elevene skal igjen lage et krysningskjema.

- e) Anta koplete gener. Hva om det hadde skjedd en overkrysning hos et av individene med genotype SsTt x SsTt. Lag krysningskjema.

Både oppgave d) og e) ble plassert i kategorien *middels* på bakgrunn av at elevene må bruke kunnskapene sine om koplete gener og overkrysning for å kunne lage krysningskjema til å besvare spørsmålene.

Den siste oppgaven som er plassert i kategorien *middels* består av tre underspørsmål der elevene blir bedt om å sammenligne replikasjon, transkripsjon og translasjon med tanke på hva som er likt og hva som skiller de fra hverandre, hvilke stoffer som er involvert og deres rolle.

12. Gi en kortfattet oversikt over replikasjon, transkripsjon og translasjon.
- Hva er likt, hva skiller disse fra hverandre?
 - Hvilke stoffer er involvert og hvilken rolle spiller de?
 - Viktig å presisere sammenhenger om hvordan ting fungerer.

På grunn av at elevene må sammenligne ulike prosesser, og at det ikke holder å bare gjengi faktakunnskaper, ble denne oppgaven plassert i kategorien *middels*. Da det ikke er noen klare grenser for hvilken kategori de ulike oppgavene bør plasseres innen, vil fremgangsmåten drøftes i kapittel 5.5 om studiens kvalitet.

4.3 Biologielevenes refleksjoner

Resultatene presenteres i form av elevenes ytringer innenfor hver av kategoriene presentert i flytdiagrammet i Figur 3. Det er altså elevenes fortellinger som står i sentrum, slik at deres opplevelse av vurderingspraksisen i biologiklassen kommer tydelig frem. Først vil analytiske påstander knyttet til elevenes refleksjoner under kategorien *kompetansemål* gjennomgås, deretter refleksjoner angående *vurderingssituasjon* og til slutt *vurdering*. Avslutningsvis oppsummerer jeg funnene.

4.3.1 Kompetansemål

Delkapittel 4.3.1 tar for seg hovedkategorien *kompetansemål* med tilhørende underkategori *læringsmål*.

Elevene uttrykker at de ikke vet hva de blir vurdert etter

Under fokusgruppeintervjuet ble elevene stilt spørsmål om de visste hvilke kompetansemål de ble vurdert etter, og om de brukte å se over dem før prøven for å orientere seg. Trine var den eneste av de fem elevene som var oppdatert på kompetansemålene. En av de øvrige elevene responderte at de hadde glemt det helt av, og ingen av de andre elevene brukte kompetansemålene aktivt når de forberedte seg til prøven. På spørsmål om elevene visste hva de blir vurdert etter, poengterte Jonas at han ikke visste det før han fikk prøven tilbake.

Jonas: Man lærer liksom bare i ettertida av prøven hva man kunne ha.. Jeg forstår jo karakterene mine først når han sier hvorfor, hva jeg må gjøre liksom..

I intervjuet med læreren kom det frem at han ofte bryter ned kompetansemålene til læringsmål muntlig underveis i undervisningsøktene, men at han ikke brukte å si direkte hva elevene ble vurdert etter. Det kan stilles spørsmål ved om elevene får med seg læringsmålene når de uttrykkes muntlig og ikke skriftlig. Det var flere som også etterlyste at læreren gjerne kunne oppgi hvor mange poeng de kunne få på hver oppgave på prøven, slik at de kan bruke det som en indikator på hva eller hvor mye som kreves. Samlet sett uttrykte elevene at de ikke visste

hva de ble vurdert etter. I diskusjonen vil jeg komme tilbake til denne trusselen for validiteten i vurderingen.

4.3.2 Vurderingssituasjon

I kapittel 4.3.2. beskrives funnene i forbindelse med hovedkategorien *vurderingssituasjon*, med tilhørende underkategorier *eksamen* og *lærebok*.

Elevene mener det er stor nivåforskjell på oppgavene på prøven og eksamen

Samtlige elever poengterer at det er stor nivåforskjell på oppgavene på prøvene de har gjennom skoleåret, og eksamensoppgavene som de eksempelvis var en del av heldagsprøven før jul. Først stilte jeg spørsmål om hva de syntes om nivået på flervalgsoppgavene på prøvene sammenlignet med heldagsprøven som inneholdt eksamensoppgaver og sammenlignet med andre eksamensoppgaver de har jobbet med.

Intervjuer: [...] Hvordan synes dere flervalgsoppgavene var på prøvene i forhold til heldagsprøven eller de eksamensoppgavene dere har sett?³

Trine: Jeg synes de på prøvene er mye lettere enn eksamensoppgavene..

Etterpå ville jeg vite hva de tenkte generelt om oppgavene. De kom tydelig frem at de mente oppgavene på heldagsprøven og eksamensoppgavene de hadde sett på krevde forståelse i større grad enn oppgavene på prøven.

Intervjuer: [...]Hva synes dere om nivået på prøvene da? Synes dere generelt det har vært enklere enn..⁴

Trine: Ja, jeg synes det er helt forskjellig lagt opp fra eksamen.. Det er liksom, jeg føler ikke at det har samme relevansen.. Det går ikke ut på de samme spørsmålene..

Intervjuer: Nei, for dere var litt inne på det med heldagsprøven at det var litt forskjell.. Så synes dere det var stor nivåforskjell?

Jonas: På de, ja..

³ [...] brukes som transkripsjonskode når deler av sitat ikke er relevant for sammenhengen.

⁴ Her er spørsmålet basert på hva elevene har sagt tidligere i intervjuet.

Trine: Ja.. Du må vise mye mer forståelse enn den puggekunnskapen..

Nora: Ja, siden på prøvene våre nå, så er det bare å omformulere det som står på den og den siden i boka på en måte.. Mens på eksamen så er det mye mer å knytte tråder og forstå.. Vise forståelse, og det er litt vanskelig når vi ikke har jobbet med det..

Trine uttrykker ytterligere en gang senere i intervjuet at prøveoppgavene er mye enklere enn heldagsprøven og eksamensoppgaver de har sett på. Elevene etterlyser også mer trening i oppgaver som ligner på eksamensoppgaver. De har brukt noen eksamensoppgaver i undervisningen i genetikk, der de har gått gjennom hvordan de skal gå frem. Samtidig er det flere av elevene som sier at de gjerne ville hatt eksamensoppgaver som lekse gjennom året, slik at de blir vant til slike oppgaver som de mener utfordrer forståelsen i mye større grad. Dette kan sammenlignes med høyere nivå i analyseverktøyet basert på Blooms taksonomi og Webbs nivåinndeling. I den sammenheng var det også interessant å høre om noen av oppgavene på prøven lignet veldig på hva de hadde gjennomgått i undervisningen. Både Beate, Jonas og Trine bekreftet av det var en klar sammenheng.

Intervjuer: [...] Var det noen av spørsmålene på prøven som var veldig lik det dere har gjennomgått?

Beate: Det med krysning. De oppgavene var veldig sånn «Lag et krysningsskjema» med det er noen som har den og den genotypen liksom, og det var ganske.. Det føler jeg vi har gått gjennom mye før.. Også var det noen som var litt mer vrien.

Jonas: Ja..

Samlet sett viser analysen angående elevenes refleksjoner rundt vanskelighetsgrad på prøveoppgavene at elevene mener de er vesentlig enklere sammenligner med eksamensoppgavene som de har sett på, og som de fikk mange av på heldagsprøven før jul. De etterlyser mer trening i oppgaver som tester deres forståelse, fordi deres erfaring tilsier at eksamen som de kan komme opp i vil kreve en annen type kunnskap enn det de blir presentert på vanligvis i vurderingssammenhenger. Jeg vil diskutere trusler for validitet som disse funnene kan ha betydning for.

Læreboken er sentral for hvordan elevene forbereder seg til skriftlige prøver

Da elevene ble spurt om hvordan de forbereder seg til vurderingssituasjon, kom det tydelig frem at de baserte seg på å lese læreboken.

Johanne: Hver gang jeg leser til prøver, så leser jeg gjennom kapitlene fra start til slutt samtidig som jeg tar notater hvis det er noe jeg synes er viktig eller noe jeg synes jeg må huske på, også skriver jeg alltid med faktaboksene for å lese gjennom dem før prøven, litt nærmere prøven.

Nora: Jeg leser gjennom også, også har jeg et slags sånt overskriftsystem. Så først lager jeg lister med alle forskjellige overskrifter i hele kapitlet, også skriver jeg om, skulle til å si nesten alt som står med mine egne ord under hver sånn undertittel, også går jeg gjennom og husker det da.

Beate: Jeg pleier egentlig å gjøre det samme til hver prøve. Starter sånn, det spørres hvor mange kapitler vi har hatt, nå har vi hatt to, da starter jeg sånn to dager før og da leser jeg liksom delkapittel og så skriver jeg notater til delkapitlet fra det jeg husker. Også pleier jeg å gå inn på bios-nettsidene å ta de der flervalgsspørsmålene fordi jeg bare vet at (lærer) liker å ta spørsmål derifra på prøven.

Ifølge elevene er det hovedsakelig læreboken de bruker aktivt når de forbereder seg til prøven. Biologilæreren bekreftet også at de i undervisningen følger læreboken ganske slavisk, og at han har inntrykk av at mange av elevene liker å lage egne notater fra læreboken. De bruker ikke å få bestemte oppgaver fra læreboken som de skal gjøre, så det er mye opp til elevene selv hvordan de vil tilegne seg lærestoffet. Samlet sett viser dette viktigheten av læreboken både til undervisning og vurdering, da elevene stort sett baserer sin kunnskap på det de leser i læreboken. I kapittel 2.2.5 tok jeg opp bruk av lærebok i undervisningen, og funnene samsvarer med Bachmanns (2005) undersøkelser.

4.3.3 Vurdering

I kapittel 4.3.3 presenteres elevenes refleksjoner under hovedkategorien *vurdering* med tilhørende underkategori *tilbakemelding*.

Elevene bruker i liten grad tilbakemeldingene på prøvene

Et viktig aspekt ved vurdering som beskrives i teoridelen er å gi tilbakemeldinger som elevene kan bruke videre i læringsprosessen. Under fokusgruppeintervjuet kom det frem at elevene i liten grad bruker tilbakemeldingene. For eksempel trekker Trine frem at hun ikke synes tilbakemeldingene er relevant for det hun skal lære videre. Beate poengterer at hun brukte tilbakemeldingene i større grad før, og at hun foretrakk de skriftlig på It's Learning.

Intervjuer: Så lurer jeg på, hvordan bruker dere de tilbakemeldingene dere har fått på prøvene tidligere? ... Eller vi kan først starte med, bruker dere dem?

Trine: Jeg glemmer dem egentlig ganske fort..

Intervjuer: Ja, så hvordan er det når du får prøven tilbake? Eller nå har det kanskje vært på It's learning eller Fronter?

Trine: Ja, de er som regel på It's Learning, og så får vi tilbake etter hvert.. Så noen ganger sender han ut en mail om hva som liksom har vært bra, hva man burde ha jobbet litt mer med, men jeg føler det er veldig spesifikt for det temaet og ikke så relevant for neste tema igjen. For at om jeg ikke har med så mye av om hvordan erteplanten blir dannet, så har jo ikke det noe å si for miljøgift i neste kapittel.. Så sånn sett, så klarer jeg ikke å trekke de trådene der videre..

Sitatet fra Trine viser at læreren gir tilbakemeldinger om hva som har vært bra, og hva som kan jobbes med videre, men hun problematiserer bruken av tilbakemeldingene videre i læringsprosessen. Hun ser ikke hvordan hun kan bruke det videre i et annet tema. Når Beate supplerer hva Trine sa, kommer det frem at læreren også har gitt muntlig tilbakemeldinger, men at det virker som om hun ikke får så mye ut av det.

Beate: Det var sånn, vi hadde [navn på lærer] i biologi 1 og da fikk vi liksom tilbakemelding på It's learning nesten hver prøve..og det var ganske greit for da følte jeg at jeg brukte det mye mer. Nå er det mer sånn muntlig tilbakemelding, så «Ja, du må drøfte mer», «ja, det vet jeg. Takk» og går og setter meg liksom.. Også er jeg veldig sær, jeg titter ikke på prøvene mine vanligvis.. Jeg liker ikke å få karakter. Da havner den mest i sekken, så det er mer sånn.. Ja.. Det er det den havner..

Jonas: Jeg kan relatere til den taktikken..

Intervjuer: Ja, så du kjenner deg igjen?

Jonas: Ja, jeg ser på den hjemme, jeg..

Nora: ..alle er redd for karakteren sin da..

Stobart (2012) poengterer at når elevene ikke bruker tilbakemeldingene fra vurderingssituasjon, vil det kunne svekke validiteten i vurderingen. Dette vil jeg komme tilbake til i kapittel 5.2.3

4.4 Oppsummering av funn

Resultatene fra analysen av hvilke kognitive ferdigheter kompetansemålene og prøveoppgavene krevde av elevene, indikerte at både kompetansemålene og prøveoppgavene for det meste krevde lavere- og middels kognitive ferdigheter av elevene. Det at prøveoppgavene ikke krevde høyere kognitive ferdigheter av elevene, bekrefter annen forskning (Bergem et al., 2015). Samtidig kan prøven omtales som å inneha innholdsvaliditet.

Ut i fra elevenes refleksjoner om vurderingspraksisen i deres biologi 2-klasse, kom det frem fire analytiske påstander som kort fortalt tar for seg:

- *Uklare læringsmål*
- *Nivåforskjell på prøve og eksamen*
- *Bruk av lærebok*
- *Liten bruk av tilbakemeldinger*

I kapittel 5 vil jeg drøfte funn og analyseverktøyet som ble brukt for å analysere kompetansemålene og prøveoppgavene. Deretter vil jeg se nærmere på de fire analytiske påstandene som er lagt frem i dette kapitlet, og knytte dem opp mot teori om trusler for validitet.

5 Diskusjon

For å besvare problemstillingen, vil jeg i dette kapitlet diskutere funnene som er presentert i analysen i kapittel 4. Forskning og teori fra kapittel 1 og 2 vil brukes for å støtte opp om eller utfordre funnene. Først vil jeg se nærmere på funnene knyttet til analysen av kompetansemålene og prøveoppgavene, samt drøfte styrker og svakheter med analyseverktøyet. Deretter drøftes funnene fra analysen av fokusgruppeintervjuet med biologielevene. Videre vil jeg bruke empiri fra studien og teori fra kapittel 2 til å utvikle spørsmål som kan supplere validitetskjeden, slik at biologilærere kan bruke det som utgangspunkt for å sikre best mulig validitet i vurderingspraksisen. Avslutningsvis vil jeg kommentere didaktiske refleksjoner før studiens kvalitet drøftes.

5.1 Hvordan kan kompetansemål og prøveoppgaver analyseres for å gi innsikt i hvilke kognitive ferdigheter de krever hos elevene?

I denne studien har jeg brukt et analyseverktøy basert på Blooms taksonomi og Webbs nivåinndeling for å analysere kompetansemål og prøveoppgaver i genetik i biologi 2. Hensikten var å undersøke hvilke kognitive ferdigheter kompetansemålene og prøveoppgavene krevde av elevene.

5.1.1 Samsvar mellom kompetansemål og prøveoppgaver

Resultatene tydet på at både kompetansemålene og prøveoppgavene stort sett forutsetter bruk av lavere og middels kognitive ferdigheter. Det at de fleste prøveoppgavene tester lavere kognitive ferdigheter, samsvarer med tidligere forskning (Boesen, 2006; Jones, Harland, Reid & Bartlett, 2009; Lord & Baviskar, 2007). Samsvaret mellom kompetansemål og prøveoppgaver kan oppsummeres som en indikasjon på at prøven har innholdsvaliditet. Det vil si at det var samme nivå på de kognitive ferdighetene som kompetansemålene krevde hos elevene, som på prøveoppgavene. Samtidig er det viktig å trekke frem at biologifaget krever at eleven har oversikt over mange ulike begreper og prosesser. Det å gjengi dette kan kanskje i utgangspunkt tenkes å regnes som lavere kognitive ferdigheter, men samtidig kan disse begrepene og prosessene i utgangspunktet være kompliserte.

Analyseverktøyet ble utviklet for å besvare det første forskningsspørsmålet, og jeg vil videre drøfte styrker og svakheter med bruk av verktøyet.

5.1.2 Styrker med analyseverktøyet

Analyseverktøyet ble utviklet med utgangspunkt i Blooms taksonomi og Webbs nivåinndeling. Årsaken til at Webbs nivåinndeling ble inkludert var at den kunne bidra med å se kompetansemålene og prøveoppgavene i en dypere kontekst enn kun ved bruk av Blooms taksonomi alene. Det vil jeg trekke frem som en styrke ved verktøyet.

Forskning har lagt vekt på at Blooms taksonomi alene kan bidra til å gjøre læreren mer bevisst på å bruke et større spekter av oppgaver som tester ulike kognitive ferdigheter på prøver og ellers i biologiundervisning (Crowe, Dirks & Wenderoth, 2008). I tillegg er denne taksonomien kjent i det norske skoleverket (Eggen & Vidnes, 2015). Med andre ord kan en kjennskap til Blooms taksonomi være et godt utgangspunkt for å kunne ta i bruk verktøyet som er brukt i denne studien. En annen styrke vil være at den kan brukes på tvers av fag. Et eksempel er at det samme verktøyet som jeg har brukt i denne studien har Baardsgard (2014) brukt i sin masteroppgave der hun blant annet har analysert kompetansemål i kjemi 2, bortsett fra at hun kombinerte Webbs nivåinndeling med Blooms reviderte taksonomi. Krathwohl (2002) vektlegger også at Blooms taksonomi kan bidra til et felles språk om læringsmål og bidra til å sikre sammenheng mellom mål, aktiviteter og vurdering. Dette mener jeg at analyseverktøyet som er benyttet i denne studien også kan bidra til.

5.1.3 Svakheter med analyseverktøyet

Når det gjelder svakheter med analyseverktøyet vil jeg trekke frem utfordringen med å analysere åpne kompetansemål. Slik jeg erfarer det, er kompetansemålene jeg har analysert i genetikk i biologi 2 relativt lukkede. Eksempelvis «gjøre greie for transkripsjon og translasjon av gen og forklare hvordan regulering av gen kan styre biologiske prosesser. Hvis kompetansemålene er åpne og vide, vil det kunne være vanskeligere å bruke analyseverktøyet til å si noe om de kognitive ferdighetene som kreves hos elevene. Et annet tilfelle vil være om kompetansemålene krever bruk av andre egenskaper enn kognitive ferdigheter, som for

eksempel praktisk arbeid, holdninger og bruk av ulike sanser. Da vil ikke dette analyseverktøyet være like relevant å bruke.

Fjørtoft (2016) har også kritisert Blooms taksonomi fordi det kan det være utfordrende å skille mellom vanskegraden på oppgaver som plasseres på ulike taksonomiske nivå. Dette er prøvd å forenkles ved å ta i bruk Webbs nivåinndeling i tillegg. Til slutt vil jeg også løfte frem synspunktene til Ritchhart, Church og Morrison (2011) om at forståelse ikke følger en lineær utvikling, men heller er en prosess som beveger seg frem og tilbake. Da vil det kanskje ikke gi mening å undersøke hvilken kognitiv ferdighet en oppgave eller et kompetansemål krever, men Ritchard et al. (2011) sier at det er viktig å tenke på konteksten og bringer i den sammenheng inn begrepene dyp og overflatisk læring. Jeg vil ikke gå videre på det her, men jeg mener det er en viktig refleksjon å bemerke.

Oppsummert kan dette analyseverktøyet brukes til å analysere kompetansemål og prøveoppgaver for å gi innsikt i hvilke kognitive ferdigheter de krever hos elevene. Jeg ser flere fordeler enn ulemper ved å ta i bruk analyseverktøyet, og jeg legger særlig vekt på at det kan bidra til en bevisstgjøring hos læreren i forhold til hva som testes og det å kunne se sammenheng mellom mål, aktiviteter og vurdering i undervisningen. Videre vil jeg drøfte de analytiske påstandene fra analysen av elevenes refleksjoner rundt vurderingspraksisen.

5.2 Hvordan kan en analyse av validitetskjeden brukes som et verktøy av læreren for å sikre best mulig validitet i vurderingspraksisen?

For å besvare forskningsspørsmålet har vurderingspraksisen blitt analysert på bakgrunn av elevenes refleksjoner. Ut i fra funnene har jeg presentert analytiske påstander som kan si noe om trusler for validiteten i vurderingen. Jeg vil drøfte funnene i lys av teori og tidligere forskning, og på bakgrunn av teori og empiri supplere validitetskjeden slik den er presentert av Sandvik og Fjørtoft (2016) med spørsmål som lærere kan stille seg for å sikre en best mulig validitet i vurderingspraksisen. De påfølgende delkapitlene deles inn etter hovedkategoriene fra Figur 3, *kompetansemål*, *vurderingssituasjon* og *vurdering*, som alle er deler av validitetskjeden.

5.2.1 Kompetansemål

Under hovedkategorien *kompetansemål* ble det identifisert ett funn knyttet til læringsmål som kan utgjøre en trussel for validitet i vurderingen.

1. *Elevene uttrykker at de ikke vet hva de blir vurdert etter*

Under fokusgruppeintervjuet kom det frem at kun én av elevene hadde kjennskap til kompetansemålene. Biologilæreren sa selv at han noen ganger brukte å bryte kompetansemålene muntlig opp til læringsmål underveis i øktene, men elevene var ikke bevisste på det under intervjuet. De visste heller ikke hvilke vurderingskriterier de ble vurdert etter, og det var ikke oppgitt antall mulige poeng på de ulike oppgavene. Disse funnene der elevene ikke kjenner til læringsmålene eller vurderingskriterier samsvarer med funnene Sandvik et al. (2014) har gjort i matematikk på en norsk videregående skole. Stobart (2012) trekker frem dette som en trussel for validiteten i vurderinger, og peker på at det er viktig at eleven vet hva som er formålet og hva en god presentasjon er. Tydelige mål og vurderingskriterier er også noe William og Thomson (2007) vektlegger. Det at elevene vet hva de skal lære og hva som forventes av dem er også et av de grunnleggende prinsippene for god vurderingspraksis i den nasjonale satsningen på vurdering for læring (Utdanningsdirektoratet, 2011).

Samtidig kan for spesifiserte læringsmål føre til fragmentert overflatelæring, så her vil det være en balansegang (Stobart, 2012). Når det gjelder kompetansemålene gitt i hovedområdet genetikk, oppfatter jeg dem som relativt spesifiserte og det kan da være en overveining for læreren om de bør gjøres om til mindre læringsmål. Uansett vil utydelige mål være en trussel for validiteten i vurderingen. Ifølge Sandvik og Fjørtoft (2016) kan en løsning være at læreren og elevene utgjør et felles tolkningsfellesskap, slik at både læreren og elevene har kjennskap til hva de vurderes etter. Slik kan også elevene få et eierskap til det som skal vurderes. Langseth (2016) hevder at en større elevinvolvering og eksplisering av læringsprosessene gjennom felles begrepsforståelse med utgangspunkt i validitetskjeden vil kunne bidra til å kunne diskutere vurdering og læring i den faglige konteksten.

5.2.2 Vurderingssituasjon

Under hovedkategorien *vurderingssituasjon* ble det identifisert to funn som kan ha betydning for validiteten i vurderingen. Videre vil jeg presentere hver analytiske påstand med påfølgende drøfting.

2. Elevene mener det er stor nivåforskjell på oppgavene på prøven og eksamen

Elevene sa ved flere anledninger at de merket det var stor forskjell på hva prøveoppgavene krevde av forståelse sammenlignet med eksamensoppgavene de hadde hatt på heldagsprøven før jul. Biologilæreren bemerket at det var viktig å gi elevene trening i eksamensoppgaver da de var annerledes enn «hva er...» og «forklar hva...», mens elevene mente at de gjerne kunne få mer trening i slike type oppgaver gjennom skoleåret. Det kan tyde på at eksamen er mindre styrende for hvordan læreren legger opp undervisningen og hvordan elevene jobber med fagstoff sammenlignet med studien til Raaheim og Hauge (2007) og rapporten fra Utdanningsdirektoratet (2015) om naturfagene. Når eksamen er styrende for undervisningen kan det være en trussel for validiteten, men i dette tilfellet er jeg usikker på hva resultatene indikerer med tanke på validitet.

Et aspekt jeg synes er viktig å trekke frem er at vurderingen som gjøres gjennom skoleåret kan regnes som et kontekstuel element i elevenes læringsomgivelser og at det kan forme elevenes læring gjennom deres oppfatninger av innhold og kognitive ferdigheter som kreves for å gjøre det bra på prøven (Hall et al., 2011). Når elevene erfarer prøver der det er tilstrekkelig å reprodusere og gjengi kunnskap, kan de selektivt velge å studere det som sammenfaller med forventningen av vurderingen (Entwistle & Entwistle, 1992). Et spørsmål som dukker opp her, er hvordan nivåforskjellen kan være så stor når både prøveoppgavene og eksamensoppgavene tar utgangspunkt i kompetansemålene innenfor fagområdet? Kanskje kan en mulig forklaring være at kompetansemålene kan tolkes forskjellig, og at biologilæreren ønsker at elevene skal få kjennskap til de grunnleggende begrepene og prosessene før de blir utfordret på et høyere nivå.

3. Læreboken er sentral for hvordan elevene forbereder seg til skriftlige prøver

Som følge av at læreboken er sentral for hvordan undervisningen legges opp, var det ikke overraskende at elevene i hovedsak brukte læreboken i forberedelser inn mot prøven. Det at læreren i stor grad bruker læreboken som utgangspunkt for undervisningen er noe som går igjen i flere studier (Bachmann, 2005; Hodgson, Rønning og Tomlinson, 2012). Sandvik og Buland (2013) viser også til at engelsklærere i stor grad støtter seg til læreboken i lokalt arbeid med læreplan på alle trinn. Det kan da være fare for at læreboken danner grunnlaget for hva som regnes som pensum, og ikke kompetansemålene. I disse tilfellene kan det utgjøre en trussel for validiteten. Ifølge Wiggins og McTighe (2005) kan ikke lærebøker erstatte tydelige kompetansemål, god vurdering og gode oppgaver, men de kan være en god ressurs å bruke i undervisningen. Det er lærerens oppgave å relatere læringsmål til elevenes kunnskaps- og erfaringsbakgrunn og samfunnet de lever i (Engseth, 2016), og det tror jeg er enklere å gjøre når læreboken ikke i for stor grad definerer pensum.

Nora sa underveis i fokusgruppeintervjuet at hun bare kunne gjengi det som sto i læreboken for å gjøre det bra på prøven, og da kan en stille spørsmål om hvilken kompetanse som egentlig testes. Momsen et al. (2013) poengterer at elever vil ha fokus på å lære seg det de må kunne til de ulike prøvene de har i løpet av skoleåret, og hvis det er tilstrekkelig for elevene å gjengi teksten fra læreboken kan det utgjøre en trussel for validiteten i vurderingen. Fjørtoft (2016) påpeker også at elever som jobber med å avdekke fagets sentrale ideer med ulike ressurser, vil trolig kunne oppnå en forståelse som er mer avansert enn de som kun bruker informasjonen fra læreboken.

5.2.3 Vurdering

Under hovedkategorien vurdering ble det identifisert ett funn som kan utgjøre en trussel for validiteten i vurderingen, som presenteres her med påfølgende drøfting.

4. Elevene bruker i liten grad tilbakemeldingene på prøvene

Elevene ble stilt spørsmål om hvordan de brukte tilbakemeldingene de fikk på prøvene, og det ble da tydelig at de brukte dem i liten grad. To av elevene sa de brukte å legge prøvene rett i sekken når de fikk de tilbake, men at de noen ganger kunne se over dem hjemme. Trine sa at

hun syntes det var vanskelig å se relevansen av tilbakemeldingene hun fikk på en prøve til neste vurderingssituasjon om et nytt tema. Det at tilbakemeldingene brukes i liten grad viser også Engseth (2016) til i engelskfaget. Viktigheten av gode tilbakemeldinger og hva som påvirker om de gir en positiv læringseffekt har det blitt gjort forskning på (Hattie & Timperley, 2009). Det er mange faktorer som kan spille inn, slik som kvalitet på tilbakemeldingen som gis og motivasjon til den lærende. Men hva når eleven ikke bruker tilbakemeldingene videre i læringsprosessen? Stobart (2012) poengterer at dette kan utgjøre en trussel for validiteten i vurderingen. Elevene var tydelige på at det var karakteren som var av betydning, og at de følte et stort press for å gjøre det best mulig. Stobart (2012) poengterer at elever kan identifisere seg med karakteren og at de sammenligner seg med de andre i klassen. Hvis elevene kan se på prøven som en del av en større læringsprosess, så vil de kanskje kunne bruke tilbakemeldingene mer aktivt?

Black og William (2009) trekker frem formativ kriteriebasert vurdering, hvor elevbesvarelsene vurderes opp mot forhåndsbestemte læringsmål og vurderingskriterier. Disse kan brukes aktivt når elevene får tilbakemeldinger, og at det kan gi lærer og elever et tolkningsfelleskap. Dette kan gi et mer transparent læringsmiljø, og kanskje kan det bidra til at elevene velger å bruke tilbakemeldingene i større grad? Trolig kan det bidra til å øke validiteten i vurderingen.

Oppsummert har jeg nå drøftet fire analytiske påstander basert på biologielevenes refleksjoner, og sett hvordan det kan utgjøre en trussel for validiteten i vurderingspraksisen i biologi 2-klassen. Videre vil jeg på bakgrunn av drøftingen legge frem spørsmål som kan supplere validitetskjeden. Hensikten er at biologilærere kan bruke det som et verktøy for å sikre best mulig validitet i vurderingspraksisen.

5.3 Validitetskjeden med supplerende spørsmål

I forrige delkapittel gikk jeg gjennom ulike trusler for validitet i vurderingspraksisen i en biologi 2-klasse. På bakgrunn av det vil jeg i Tabell 4 presentere spørsmål knyttet til *kompetansemål, vurderingssituasjoner og vurdering*, som alle er deler av validitetskjeden.

Tabell 4. Trinn i validitetskjeden supplert med spørsmål som kan bidra til at biologilæreren sikrer best mulig validitet i vurderingspraksisen.

Validitetskjeden	Supplerende spørsmål
Kompetansemål	Trenger jeg bryte ned kompetansemålene til læringsmål for at det skal bli mer forståelig for elevene? Kjenner elevene til kompetansemålene/læringsmålene? Har vi et tolkningsfellesskap? Hvilke kognitive ferdigheter krever kompetansemålene av elevene?
Vurderingssituasjoner	Er det samsvar mellom kompetansemål/læringsmål, undervisning og vurderingssituasjon? Hvordan kan elevene vise at han/hun har nådd målet? Får elevene mulighet til å vise sin kompetanse i ulike vurderingssituasjoner? Er vurderingssituasjon designet slik at elevene får vist ønsket kompetanse/konstrukt? Hvilke kognitive ferdigheter dekker vurderingsoppgavene sammenlignet med kompetansemålene?
Vurdering	Brukes vurderingskriteriene som utgangspunkt for vurderingen, og kjenner elevene til dem? Hvilken tilbakemelding gis, og bruker elevene tilbakemeldingene videre i læringsprosessen? Hvordan bør undervisningen legges opp videre på bakgrunn av innsamlet informasjon om elevenes læring?

Flere av spørsmålene presentert i Tabell 4 som har til hensikt å supplere validitetskjeden tar utgangspunkt i at biologielevene og biologilæreren utgjør et tolkningsfellesskap. Dette er noe som vektlegges i litteratur når det stilles spørsmål ved hva som kan bidra til å øke validiteten i vurderingspraksisen (Sandvik og Fjørtoft, 2016). Både med tanke på at lærer-elev har samme forståelse for hva som forventes, og at lærerkollegiet kan utgjøre et tolkningsfellesskap.

Kompetansemål er nært knyttet opp til læringsmål, vurderingskriterier og kjennetegn på måloppnåelse, men jeg har kun tatt utgangspunkt i kompetansemål fordi elevene jeg intervjuet ikke hadde kjennskap til noen av de andre elementene. Med litt omformulering vil tabellen også kunne brukes som et felles utgangspunkt for lærer og elever for diskusjon om vurdering og læring. Avslutningsvis vil jeg poengtere at et viktig aspekt innen validitet er å tenke på

konsekvenser, så et naturlig spørsmål biologilæreren kan stille seg for å se alt under ett er hvilke konsekvenser vurderingspraksisen kan ha for elevenes læringsprosess.

5.4 Didaktiske refleksjoner

Når det gjelder det didaktiske perspektivet på denne oppgaven, så er det mye fra denne studien som jeg kan ta med meg videre i lærergjernen. Det største bidraget vil være en bevisstgjøring på hvilke trusler for validitet jeg kan møte i vurderingsarbeidet, og hva jeg kan gjøre for å sikre best mulig validitet i vurderingene. Validitetskjeden med de supplerende spørsmålene er noe jeg vil ta med meg videre, som et verktøy for å skape en sammenheng mellom kompetansemål, undervisning og vurdering for elevene. Når det gjelder de ulike truslene for validitet i vurderingspraksisen, er det spesielt viktigheten av at jeg som biologilærer og mine kommende elever må ha en felles forståelse for kompetansemål, læringsmål og vurderingskriterier som har gjort særlig inntrykk.

Analyseverktøyet jeg har brukt for å analysere kompetansemål og prøveoppgaver, har bidratt til å se viktigheten av hva som testes i forhold til kompetansemålene. Et annet aspekt jeg har reflektert rundt, er hvor viktig prøveoppgavene kan være for hva elevene velger å rette oppmerksomheten sin mot i læringsprosessen.

Det har vært lagt lite oppmerksomhet rundt validitet i vurderingen i løpet av lektorutdanningen, og derfor har det vært en berikelse å få mer kunnskap innen dette forskningsfeltet. Vurdering vil være en stor del av læreryrket, og arbeidet med denne studien har bidratt til å gi meg en større innsikt i vurdering generelt og trusler for validitet i vurdering spesielt. Videre vil jeg drøfte studiens kvalitet.

5.5 Studiens kvalitet

Utover det jeg har kommentert om metodekvalitet i kapittel 3.5 vil jeg utdype refleksjoner jeg har gjort meg under arbeidet med denne studien. Jeg vil kommentere analyseverktøy, utvalg, intervju og andre datakilder som kunne vært brukt.

Når det gjelder bruk av analyseverktøyet basert på Blooms taksonomi og Webbs nivåinndeling, vil jeg si at det er en styrke at jeg i en pilotoppgave forrige semester analyserte oppgaver gitt på tre ulike prøver i biologi 1. Analyseverktøyet jeg benyttet da var basert på Blooms taksonomi og Doyle. Erfaringen jeg gjorde meg da førte til at jeg valgte å kombinere Blooms taksonomi med Webbs nivåinndeling, på grunn av at sistnevnte tar større hensyn til kontekst. Det å plassere kompetansemål og prøveoppgaver i ulike grupper etter hvilke kognitive ferdigheter de krever hos elevene har jeg opplevd som utfordrende. En svakhet her er at jeg ikke har vært tilstede i undervisningen, slik at jeg ikke vet nøyaktig hva som har blitt gjort i forkant av prøven. Et viktig poeng er om biologilæreren har gjennomgått helt like oppgaver som elevene fikk på prøven, og at elevene da blir testet i hukommelse fremfor forståelse. Dermed kan det hende jeg har plassert oppgaver innenfor middels kognitive ferdigheter, når de egentlig krever reproduksjon fra undervisningen.

Biologielevene som deltok i fokusgruppeintervjuet om vurderingspraksisen var et tilgjengelighetsutvalg, det vil si at det var de fem elevene i biologiklassen som meldte seg til å delta i intervjuet. Et sentralt spørsmål vil da være om jeg hadde fått samme resultat om jeg intervjuet andre elever i klassen. Det ville vært en styrke å gjennomføre fokusgruppeintervju med to grupper, slik at refleksjonene kunne blitt sammenlignet. Et annet moment er hvordan elevene kan ha påvirket hverandres svar under intervjuet. Det hadde ikke vært tilfelle ved individuelle intervju, men samtidig ser jeg styrken med fokusgruppeintervju som kan by på rike interaksjoner der elevene spiller videre på hverandres fortellinger. En fare ved fokusgruppeintervju er hvis enkelte deltakere ikke slipper til med sine meninger, men jeg erfarte at alle var delaktige.

Utvalget av prøven med oppgaver som har blitt analysert var en av flere prøver elevene hadde hatt gjennom skoleåret. Jeg vil stille spørsmålstegn ved om denne prøven er representativ for de skriftlige vurderingssituasjonene elevene møter gjennom skoleåret i denne biologiklassen.

Både biologilæreren og elevene ga uttrykk for at det var slike prøver de brukte å ha. Biologilæreren sendte meg flere prøver som han hadde utformet, og uten å ha analysert disse oppgavene fikk jeg et inntrykk av at de var bygget opp på lik måte.

I ettertid av intervjusituasjon, når jeg har fordypet meg ytterligere i vurdering og validitet, ser jeg at jeg kunne valgt litt andre vinklinger på spørsmålene i intervjuguiden som kanskje kunne fått belyst flere trusler for validitet i vurderingspraksisen. Før intervjusituasjonen var jeg bevisst på at jeg ikke skulle stille ledende spørsmål, men jeg erfarte under transkripsjonen av fokusgruppeintervjuet at jeg hadde stilt noen spørsmål som var ledende. Dette belyser også min rolle i studien, og hvordan jeg kan ha innvirkning på resultatene. Hvis min rolle blir for fremtredende, vil det kunne medføre at andre forskere ikke hadde kommet frem til samme resultat. Jeg har gjort mange tolkninger i studien, men samtidig har jeg hele veien prøvd å være bevisst på at det begrunnes med teori.

Jeg tok et aktivt valg om at det var elevenes fortellinger som skulle være sentral i studien, og derfor valgte jeg ikke å observere undervisningen. I ettertid ser jeg at jeg med fordel kunne ha intervjuet elevene først, og deretter observert undervisningen for å ha en dypere innsikt i vurderingspraksisen i denne biologi 2-klassen. Dette ville også sikret flere datakilder, som ville vært en kvalitet for studien.

6 Avslutning

I denne studien har jeg undersøkt hvilke trusler for validitet biologilærere kan møte på i ulike faser av sin vurderingspraksis, og hvordan disse truslene kan identifiseres. Avslutningsvis vil jeg komme med avsluttende refleksjoner og besvare forskningsspørsmålene, samt komme med en konklusjon på problemstillingen og kommentere veien videre.

6.1 Avsluttende refleksjoner

På bakgrunn av egne erfaringer og tidligere forskning var det ikke noen av funnene i studien som var overraskende, men det forskningen min kan bidra med er en bevisstgjøring hos biologilærere om hvilke trusler for validitet de kan møte på i vurderingspraksisen. Samt en bevisstgjøring på hva man krever hos elevene i vurderingssituasjoner sammenlignet med kompetansemålene. I tillegg til at spørsmålene som jeg har knyttet til kompetansemål, vurderingssituasjon og vurdering i validitetskjeden kan være et bidrag til biologilærere for å kunne identifisere trusler for validitet i vurderingspraksisen. Dette er også noe jeg selv vil ta med meg videre i lærergjeringen.

Samtidig vil jeg bemerke at studier viser at det finnes mange ulike vurderingspraksiser i den norske skolen (Sandvik et al., 2014), og at jeg kun har forsket på en av mange. For at det skal skje en utvikling i skolens vurderingspraksis, viser forskning at det ikke er opp til den enkelte lærer, men at man må jobbe mot et felles mål i lærerkollegiet der også skolens ledelse er støttende til å utvikle vurderingskompetansen (Sandvik et al., 2014). Dette poengterer også viktigheten av SKUV som bidrar med en skolebasert kompetanseutvikling i vurdering.

6.2 Svar på forskningsspørsmål

1. Hvordan kan kompetansemål og prøveoppgaver analyseres for å gi innsikt i hvilke kognitive ferdigheter de krever av elevene?

Kompetansemålene i læreplanen i biologi 2 er førende for hvilken kompetanse elevene skal oppnå, og en del av vurderingspraksisen til en biologilærer er å utforme prøver hvor elevsvarene brukes som bevis for elevenes kompetanse i faget. For å finne ut hvilke kognitive ferdigheter kompetansemålene og prøveoppgavene tok utgangspunkt i, ble de analysert ved hjelp av et verktøy basert på Blooms taksonomi og Webbs nivåinndeling. Årsaken til at Blooms taksonomi ble kombinert med Webbs nivåinndeling, er at sistnevnte tar hensyn til

kontekst. Funn fra analysen var at både kompetansemålene og prøveoppgavene indikerte bruk av lavere og middels kognitive ferdigheter, og dermed så ut til å samsvare. Dette kan oppsummeres som en indikasjon på at prøven har innholdsvaliditet med tanke på de kognitive ferdighetene som kreves.

2. Hvordan kan validitetskjeden og identifisering av trusler for validitet brukes som verktøy av biologilærere for å sikre best mulig validitet i vurderingspraksisen?

Validitetskjeden slik den er presentert av Sandvik og Fjørtoft (2016) består av syv trinn som kan bidra til å sikre validitet i lærerens arbeid fra operasjonalisering av kompetansemål, til arbeidsoppgaver og vurdering. Tre av trinnene, *kompetansemål, vurderings situasjon og vurdering*, har blitt analysert i denne studien på bakgrunn av elevenes refleksjoner om vurderingspraksisen i biologi 2-klassen. På bakgrunn av dette supplerte jeg validitetskjeden med spørsmål som kan fungere som et verktøy for biologilæreren for å sikre best mulig validitet i vurderingspraksisen.

6.3 Konklusjon

Basert på funnene i studien vil jeg besvare problemstillingen: *Hvilke trusler for validitet kan biologilærere møte på i ulike faser av sin vurderingspraksis, og hvordan kan disse identifiseres?*

Da jeg skulle undersøke hvilke trusler for validitet biologilærere kan møte på i ulike faser av vurderingspraksisen, ønsket jeg å analysere oppgaver gitt på en prøve i biologi 2 for å se hvilke kognitive ferdigheter de krevde av elevene. Jeg analyserte kompetansemålene og oppgavene på prøven ved hjelp av analyseverktøyet basert på Blooms taksonomi og Webbs nivådeling. Funnene indikerte at både kompetansemålene og prøveoppgavene krevde bruk av lav- og middels kognitive ferdigheter hos elevene. På bakgrunn av det kunne jeg si at prøven var valid med tanke på innhold. Med andre ord kan prøveinnholdet være en trussel for validitet i vurderingspraksisen, og analyseverktøyet basert på Blooms taksonomi og Webbs nivåinndeling kan brukes for å identifisere den eventuelle trusselen for validiteten.

For å se prøven som en del av en større kontekst, i dette tilfellet biologilærerens vurderingspraksis, intervjuet jeg biologielever om hvordan de opplevde vurderingspraksisen i biologi 2-klassen. På bakgrunn av analysen av biologielevenes refleksjoner utviklet jeg fire analytiske påstander som alle sier noe om trusler for validiteten i vurderingspraksisen.

Oppsummering av analytiske påstander basert på elevenes refleksjoner:

1. *Elevene uttrykker at de ikke vet hva de blir vurdert etter*
2. *Elevene mener det er stor nivåforskjell på oppgavene på prøven og eksamen*
3. *Læreboken er sentral for hvordan elevene forbereder seg til skriftlige prøver*
4. *Elevene bruker i liten grad tilbakemeldingene på prøvene*

Jeg brukte funnene til å supplere deler av validitetskjeden med et sett av spørsmål, slik at biologilærere kan bruke det som et utgangspunkt til å identifisere trusler for validitet i vurderingspraksisen.

6.4 Veien videre

I denne studien ble tre av syv trinn i validitetskjeden analysert på bakgrunn av biologielevenes refleksjoner om vurderingspraksisen i deres klasse. Det var *kompetansemål*, *vurderingssituasjon* og *vurdering*. Ut i fra funnene ble det identifisert trusler for validitet, og det ble utarbeidet spørsmål til hvert trinn for å supplere validitetskjeden slik at biologilærere kan bruke det som et verktøy for å sikre best mulig validitet i vurderingspraksisen. Et forslag til videre studier kan være å analysere de fire gjenstående trinnene, og på bakgrunn av empiri utvikle spørsmål til alle trinnene for å få et fullstendig verktøy.

En annen vinkling på videre forskning kan være å gjøre en intervensjonsstudie på bakgrunn av empirien fra denne studien. Det kunne være interessant og observert vurderingspraksisen til en biologilærer over en periode, og deretter gitt læreren validitetskjeden med de supplerende spørsmålene for å bruke det som et verktøy videre i vurderingspraksisen. Forskeren kunne da observert om det førte til endringer av trusler for validitet i vurderingspraksisen til biologilæreren.

7 Litteraturliste

- Anderson, L.W. & Krathwohl, D.R. (red.). (2001). *A taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Angell, C., Bungum, B., Henriksen, E.K., Kolstø, S.D., Persson, J., & Renstrøm, R. (2016). *Fysikkdidaktikk*. Kristiansand: Cappelen Damm Akademisk
- Bergem, O. K., Goodchild, S., Henriksen, E. K., Kolstø, S. D., Nortvedt, G. A. & Reikerås, E. K. L. (2015). *Realfag: Relevante-engasjerende-attraktive-lærerrike. Rapport fra ekspertgruppa for realfagene*. Hentet 02.12.18 fra https://www.regjeringen.no/globalassets/upload/kd/vedlegg/rapporter/rapport_fra_ekspertgruppa_for_realfagene.pdf
- Berger, P.L. & Luckmann, T. (1991). *The Social Construction of Reality*. London: Clays ltd.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-74.
- Black, P. & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. & Serret, N. (2010). Validity in teachers' summative assessment. *Assessment in education: Principles, Policy & Practice*, 17(2), 215-232. <https://doi.org/10.1080/09695941003696016>
- Black, P., Harrison, C., Lee, C., Marshall, B. & Dylan, W. (2003). *Assessment for learning. Putting in into practice*. Berkshire: Open University Press.
- Blooms, B. (1956). *Taxonomy of Educational Objectives*. New York: Longmans, Green and co.
- Boesen, J. (2006). *Assessing mathematical creativity*. Doctoral dissertation, Umeå University, Umeå.
- Bonnier, S.M. (2013). Validity in Classroom Assessment. Purposes, Properties and Principles. I: McMillian, J.C. (red.). *SAGE Handbook of Research on Classroom Assessment*, 87-106. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage Publications.
- Brookheart, S.M. (2014). *How to design questions and tasks to assess student thinking*. Alexandria: Ascd.
- Baardsgaard, M. (2014). *Innhold i kjemifaget: Hovedområdenes omfang i læreplan, lærebøker og på skriftlig eksamen i kjemi 2*. NTNU: Trondheim
- Christoffersen, L. & Johannessen, A. (2012). *Forskningsmetoder for lærerutdanningene*. Oslo: Abstrakt forlag.
- Crooks, T. J., Kane, M.T. & Cohen, A.S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education*, 3(3), 265-285.
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Science Education*, 7(4), 368-381. doi: 10.1187/cbe.08-05-0024.
- Dale, L.E. & Wærness, J.I. (2006). *Vurdering og læring i en elevaktiv skole*. Oslo: Universitetsforlaget.

- Doyle, W. (1983). Academic work. *Review of educational research*, 53(2), 159-199. doi: 10.2307/1170383
- Engh, K. R. (2011). *Vurdering for læring i skolen: på vei mot en bærekraftig vurderingskultur*. Kristiansand: Høyskoleforlaget.
- Entwistle, A. & Entwistle, N. (1992). Experiences of Understanding in Revising for Degree Examinations. *Learning and Instruction*, 2(1), 1-22. doi: [https://doi.org/10.1016/0959-4752\(92\)90002-4](https://doi.org/10.1016/0959-4752(92)90002-4)
- Eggen, P. O. & Vidnes, B. (2015). Vurdering i biologi. I: van Marion, P. og Strømme, A (red), *Biologididaktikk* (2. utg.), 240-241. Oslo: Cappelen Damm.
- Fjørtoft, H. (2016). *Effektiv planlegging og vurdering*. (2. utg.). Bergen: Fagbokforlaget.
- Goodlad, J. I. (1979). *Curriculum inquiry: the study of curriculum practice*. New York: McGraw-Hill.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational communication and technology*, 29(2), 75-91.
- Hall, K.L., Watkins, J.E., Coffey, J.E., Cooke, T.J, & Redish, E.F. (2011). *Examining the impact of students expectations on undergraduate biology education reform*. Artikkel presentert på «American Research Association National Meeting», april 2011 i New Orleans, La. Hentet 10. april 2018 fra: <http://umdberg.pbworks.com/f/Hall%202011%20AERA%20paper.pdf>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487
- Hirst, P. & Peters, R.S. (1970). *The logic of education*. London: Routledge and Keagan Paul.
- Hodgson, J., Rønning, W. & Tomlinson, P. (2012). *Sammenhengen mellom undervisning og læring. En studie av læreres praksis og deres tenkning under kunnskapsløftet, sluttrapport*. Hentet 29.03.18 fra: http://www.nordlandsforskning.no/getfile.php/132403-1412587174/Dokumenter/Rapporter/2012/Rapport_04_2012.pdf
- Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.
- Husén, T. (1988). Research paradigms in education. *Interchange*, 19(1), 17-20.
- Imsen, G. (2016). *Lærerenes verden: innføring i generell didaktikk* (5. utg.) Oslo: Universitetsforlaget.
- Jacobsen, D. I. (2005). *Hvordan gjennomføre undersøkelser? – Innføring i samfunnsvitenskapelig metode* (2. utg.). Kristiansand: Høyskoleforlaget.
- Jones, K.O, Harland, J., Reid, J. M. V., & Bartlett, R. (2009). Relationship between examination questions and Bloom's taxonomy. *Frontiers in education conference*, 39. IEEE. doi: 10.1109/FIE.2009.5350598
- Kane, M.T. (2016). Explicating validity, *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211.
- Krathwohl, D.R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), 212-218. doi: http://doi.org/10.1207/s15430421tip4104_2
- Langseth, I. (2016). Elevinvolvering i Engelskfaget?. I: Fjørtoft, H. og Sandvik, L.V. (red.), *Vurderingskompetanse i skolen*, 139-154. Oslo: Universitetsforlaget.

- Lord, T. & Baviskar, S. (2007). Moving students from information recitation to information. *Journal of College Science Teaching*, 36(5), 40-45.
- Malterud, K. (2012). *Fokusgrupper som forskningsmetode for medisin og helsefag*. Oslo: Universitetsforlaget.
- Marzano, R.J. & Kendall, J.S. (2007). *The new taxonomy of educational objectives* (2. utg.). Thousand Oaks, CA: Corwin Press.
- Messick, S. (1989). Validity. I R.L. Linn (red.), *Educational measurement* (3. utg.), 13-103 New York: Macmillan.
- Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments, *Educational Researcher*, 23(2), 13-23.
- Messick, S. (1995). Validity of Psychological Assessment: Validation of Interferences From Persons' Responses and Performance as Scientific Inquiry Into Score Meaning, *American Psychologist*, 50(9), 741-749.
- Momsen, J., Offerdal, E., Kryjevskaja, M., Montplaisir, L., Anderson, E. & Grosz, N. (2013). Using assessment to investigate and compare the nature of learning in undergraduate science courses. *CBE – Life Sciences Education*, 12(2), 239-249.
- NESH. (2016). *Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, jus og teologi*. Hentet 10.01.2018 fra: <https://www.etikkom.no/forskningsetiske-retningslinjer/Samfunnsvitenskap-jus-og-humaniora/>
- Pellegrino, J.W., Chudowsky, N., & Glaser, R. (red.), (2001). *Knowing what students know. The Science and design of educational assessment*. Washington, DC: National Academy.
- Ritchhart R., Church, M. & Morrison, K. (2011). *Making thinking visible: How to Promote Engagement, Understanding and Independence for All Learners*. San Francisco, CA: Jossey-Bass.
- Raaheim, A. og Hauge, H. (2007). Eksamen og eksamensangst. I: Tveit, S. (red.). *Elevvurdering i skolen – grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Robson, C. & McCartan, K. (2016). *Real world research: a resource for users of social research methods in applied settings* (4. utg.). Chichester: Wiley
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 18, 119-144.
- Sandvik, L.V. & Buland T. (red.) (2013). *Vurdering i skolen. Operasjonaliseringer og praksiser. Delrapport 2 fra prosjektet «Forskning på individuell vurdering i skolen» (FIVIS)*, Trondheim: NTNU Program for lærerutdanning og SINTEF.
- Sandvik, L.V. & Fjørtoft, H. (2016). Å utvikle tolkningsfellesskap i skolen: et verktøy og et kart for vurdering. I: Fjørtoft, H. og Sandvik, L.V. (red.), *Vurderingskompetanse i skolen*, 42-58. Oslo: Universitetsforlaget.
- Sandvik, L.V., Engvik, G., Fjørtoft, H., Langseth, I.D., Aaslid, B.E., Mordal, S. & Buland, T. (2012). *Vurdering i skolen – intensjoner og forståelser. Delrapport 1 fra prosjektet «Forskning på individuell vurdering i skolen» (FIVIS)*. Trondheim: NTNU, Program for lærerutdanning og SINTEF.

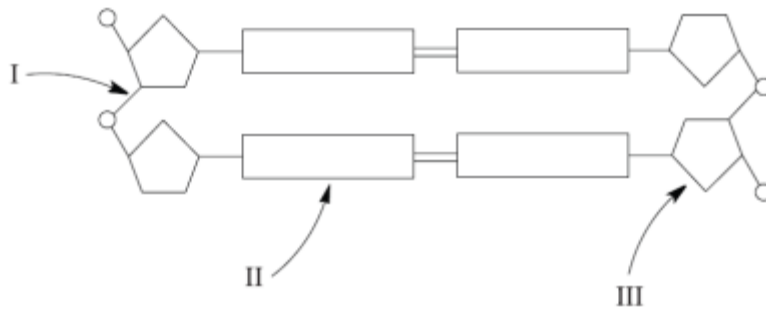
- Stobart, G. (2009). Determining validity in national curriculum assessments, *Educational Research*, 52(2), 161-179.
- Stobart, G. (2012). Validity in Formative Assessment. I: Gardner, J. (red.). *Assessment and Learning: An Introduction*. London: Sage.
- Strømme, A. (2015). Hva er egentlig biologi?. I: van Marion, P. og Strømme, A (red), *Biologididaktikk* (2. utg.), 18-38. Oslo: Cappelen Damm.
- Taras, M. (2005). Assessment: Summative and Formative: Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4), 446-478.
- Thronsen, I., Hopfenbeck, T.N., Lie, S. & Dale, E.L. (2009). *Bedre vurdering for læring. Rapport fra «Evaluering av modeller for kjennetegn på måloppnåelse i fag»*. Hentet 03.03.18 fra: https://www.udir.no/globalassets/upload/forskning/5/bedre_vurderingspraksis_ils_rapport.pdf
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning, *Assessment in Education*, 14(3), 281-294.
- Utdanningsdirektoratet (2011). *Nasjonal satsing på vurdering for læring (2010-2014)*. Oslo: Utdanningsdirektoratet.
- Utdanningsdirektoratet (2013). Læreplan i biologi – programfag i utdanningsprogram for studiespesialisering. Hentet 20. januar fra Utdanningsdirektoratet: <https://www.udir.no/kl06/BIO1-01/Hele/Kompetansemal/biologi-2>
- Utdanningsdirektoratet (2015). Naturfagene i norsk skole anno 2015 – Faggjennomgang av naturfagene – rapport fra eksternt arbeidsgruppe oppnevnt av Utdanningsdirektoratet. Hentet 03.05.18 fra: <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/naturfagene-i-norsk-skole-anno-2015/>
- Utdanningsdirektoratet (2016). *Å forstå kompetanse*. Hentet 01.03.18, fra <https://www.udir.no/laring-og-trivsel/lareplanverket/forsta-kompetanse/>
- Webb, N.L. (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph, 6. Madison: University of Wisconsin, Wisconsin Center and Education.
- Wiggins, G. & McTighe, J. (2005). *Understanding by Design*. New Jersey: Pearson Merrill Prentice Hall.
- William, D. (2011). What is assessment for learning?. *Studies in Educational Evaluation*, 37, 3-14
- Yin, R.K. (2014). *Case study research design and methods* (5. utg.). Thousand Oaks: Sage publications.

Vedlegg A: Prøve gitt i genetikk i biologi 2

1. En organisme med to identiske alleler for en egenskap blir kalt
 - homozygot
 - dominant
 - selvpollinerende
 - heterozygot
2. Når en erteplante med gule erter (frø) blir krysset med en plante som har grønne erter, får alle avkommene i F1 gule erter. Dette betyr at allelet som gir gul farge er
 - tilstede i begge foreldreplantene
 - heterozygot
 - dominant
 - recessivt
3. Hvilken krysning er eksempel på en testkrysning?
 - $TT \times Tt$
 - $Aa \times Aa$
 - $Dd \times dd$
 - $Pp \times PP$
4. En krysning mellom en purpurfarget petunia og en hvit petunia ga 100 avkom som alle var blå. Dette er et eksempel på:
 - dominant nedarving
 - intermediær arv
 - kodominans
 - kjønnsbundet nedarving
5. En krysning mellom en purpurfarget petunia og en hvit petunia ga 100 avkom som alle var blå. Når to blå petuniaer krysses med hverandre blir fordelingen i avkommet:
 - 75 % hvite, 25 % purpur
 - 75 % purpur, 25 % hvite
 - 100 % blå
 - 25 % purpur, 50 % blå, 25 % hvite
6. En mann med blodtype A får barn med en kvinne med blodtype B. Hvordan er det mulig at de får barn med blodtype O?
 - Begge foreldrene er homozygote for blodtypen sin.
 - Begge foreldrene er heterozygote for blodtypen sin.
 - Faren er homozygot for blodtype A, og moren er heterozygot for blodtype B.
 - Faren er heterozygot for blodtype A, og moren er homozygot for blodtype B
7. Figuren viser en del av et DNA-molekyl.
Hva representerer I, II og III? Sett ring rundt riktig «gruppe»
 - I - hydrogenbinding II – base III - deoxyribose
 - I - hydrogenbinding II – deoxyribose III - fosfatgruppe
 - I - elektronparbinding II – base III - deoxyribose
 - I - elektronparbinding II – deoxyribose III - fosfatgruppe

I - peptidbinding II - base III - deoxyribose

8. Ett av nukleotidene som finnes i DNA består av adenin – ribose – fosfat



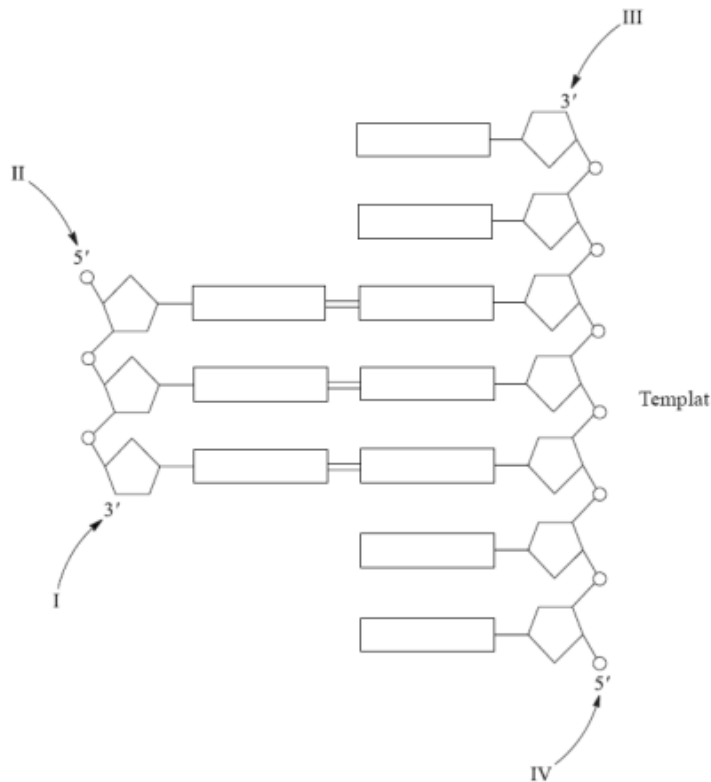
- uracil – deoxyribose – fosfat
- tymin – ribose – fosfat
- guanin – deoxyribose – fosfat
- cytosin – ribose - fosfat

9. Hva av dette er essensielt for å bevare DNA-sekvensen ved replikasjon?

- Ett gen koder for ett polypeptid
- Komplementær baseparing
- Substrater er spesifikke for enzymer
- Den genetiske koden er universell
- Fotofosforylering

10. Figuren under viser en del av et DNA-molekyl som er i ferd med å bli kopiert.

- I og II
- I
- II
- III
- IV



Skriveoppgaver på eget ark

11. Pelsfarge hos rødrev bestemmes av to gener. Begge genene må ha minst ett dominant allel for at den røde pelsen skal komme til uttrykk, hvis ikke får man sølvfarget pelsfarge.

- Dersom du krysser to sølvrever med genotypene $ssTT \times SStt$ hva blir da utfallet? Vis kryssning, samt fenotype-/genotypefrekvens.
- To dobbelt heterozygote rødrever $SsTt \times SsTt$ krysses. Lag kryssningsskjema og vis mulige utfall.
- For å bli rødrev må du ha minst en dominant av begge genene. Hvis ikke blir pelsen sølvfarget. Hvor stor er sjansen for å få sølvrever i denne kryssningen? Lag en oversikt over fenotype- og genotypefrekvens.
- Hvordan ville det sett ut om genene var koplet og to heterozygote med genotyper $SsTt \times SsTt$ ble krysset?
- Anta koplete gener. Hva om det hadde skjedd en overkryssning hos et av individene med genotype $SsTt \times SsTt$. Lag kryssningsskjema.

12. Gi en kortfattet oversikt over replikasjon, transkripsjon og translasjon.

Hva er likt, hva skiller disse fra hverandre?

Hvilke stoffer er involvert og hvilken rolle spiller de?

Viktig å presisere sammenhenger om hvordan ting fungerer.

13. Gjør rede for hovedforskjellene mellom mitose og meiose. Du trenger ikke å beskrive hver delingsfase i detalj.

Vedlegg B: Intervjuguide biologielever

Tema 1: Forberedelse til prøver

Hvilke forberedelser gjorde dere før prøven?

Hva er eventuelt likt/ulikt som tidligere?

Vet dere hvilke kompetansemål i læreplanen til biologi 2 dere blir testet i?

Bruker dere eventuelt å se over disse som forberedelse?

Vet dere hva dere blir vurdert etter?

Hvordan er det eventuelt blitt kommunisert til dere?

Vurderingskriterier, vet dere hva som kreves av svar på oppgavene?

Tema 2: Undervisningen

Hvordan har undervisningen vært lagt opp i forkant av prøven dere hadde i går?

Hvis dere tenker tilbake på undervisningen, hva tenker dere blir verdsatt?

Overflatekunnskap eller dypere forståelse?

Var noen av spørsmålene veldig like dem dere har gjennomgått?

Kunne dere bruke lik fremgangsmåte, og bare bytte ut tall for eksempel?

Tema 3: Oppgavene på prøven

Hvordan er nivået på oppgavene på prøven sammenlignet med det dere gjør i undervisningen?

Føler dere at dere får vist deres kompetanse?

Si litt mer om det er god variasjon eller samsvar med det dere forventer.

Hvilken type kunnskap synes dere at dere blir testet i, er det mye faktakunnskaper eller oppgaver der dere skal se ting i en større sammenheng?

Hva synes dere om flervalgsoppgavene? Fordeler, ulemper

Hva tenker dere at flervalgsoppgavene tester? Faktakunnskaper eller..

Ville dere eventuelt hatt flere eller færre av dem? Hvorfor?

Er oppgavene formulert slik at dere vet hva som skal gjøres?

Tema 4: Læringsprosess

Opplever dere at dere kan bruke de ulike begrepene i en større sammenheng, eller blir det pugg som å huske ett og ett begrep tatt ut av kontekst?

Tenker dere at det dere lærer er nyttig for fremtiden deres? Utdyp gjerne hva.

Tema 5: Tilbakemeldinger

Hvordan bruker dere tilbakemeldingene dere får på prøvene?

Tema 6: Motivasjon

Hva er motivasjon deres i faget?

Tema 7: Avslutning

En oppsummering. Er det noe dere vil legge til?

Vedlegg C: Intervjuguide biologilærer

Kort presentasjon: utdanningsbakgrunn, fag og år i skolen.

Tema 1: Vurderingspraksis

Hvilke typer vurderinger bruker du?

Hva vektlegger du i vurderingen av biologi 2?

Bruker du noen verktøy som hjelp i vurderingsarbeidet (hvordan)

Hvordan bruker du teknologi i vurderingsarbeidet?

Har du endret vurderingspraksis i løpet av din tid som lærer? Evt. hva?

Hvordan har du lært vurdering?

Hvorfor vurderer du slik du gjør?

Tema 2: Undervisning

Kan du beskrive hvordan undervisningen vanligvis er bygd opp?

Hvor mye brukes læreboken i undervisningen?

Tema 3: Skriftlige prøver

Kan du fortelle litt rundt arbeidet når du utformer skriftlige prøver i biologi 2?

Hva tar du hensyn til når du utformer spørsmålene på prøvene?

Hvorfor bruker du flervalgsoppgaver?

Kan du fortelle litt om utformingen av flervalgsoppgavene, eks. svaralternativene.

Har de ulike temaene mye å si for dybden i spørsmålsformuleringene? Forklar.

Har spørsmålene på prøvene blitt stilt på prøver tidligere år, hvorfor/hvorfor ikke?

Hvordan vil du si oppgavene du bruker på prøvene i biologi 2 er sammenlignet med de i biologi 1 og naturfag (fag som er muntlig-praktisk)?

Hvilken betydning har eksamen i biologi 2 for utformingen av prøvene?

Tema 4: Elevenes læringsprosess

Hvordan forventer du at elevene forbereder seg til prøvene?

Hvordan bruker du dine vurderingsresultater videre i undervisningen?

Hvordan påvirker vurderingspraksisen din måte å jobbe som lærer?

Hvordan påvirker vurdering elevenes læringsprosess?

Tema 5: Kollegier/ledelse/SKUV

Er vurdering noe som diskuteres mellom lærerne?

Har skolen en kollektiv vurderingspraksis eller mange individuelle praksiser?

Er skoleledelsen en pådriver for å utvikle skolens vurderingspraksis?

Har SKUV-prosjektet hatt noe å si for din vurderingspraksis, isåfall hva?

Tema 6: Avslutning

En oppsummering, er det noe du vil legge til?

Vedlegg D: Bekreftelse på NSD-søknad



John Alexander Strømme

7491 TRONDHEIM

Vår dato: 23.10.2017

Vår ref: 56097 / 3 / STM

Deres dato:

Deres ref:

Vurdering fra NSD Personvernombudet for forskning § 31

Personvernombudet for forskning viser til meldeskjema mottatt 21.09.2017 for prosjektet:

<i>56097</i>	<i>Etterarbeid med skriftlige prøver i biologi 1 og 2; elevslæring</i>
<i>Behandlingsansvarlig</i>	<i>NTNU, ved institusjonens øverste leder</i>
<i>Daglig ansvarlig</i>	<i>John Alexander Strømme</i>
<i>Student</i>	<i>Susanne Dyrstad</i>

Vurdering

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

Vilkår for vår anbefaling

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

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

Vi forutsetter at du ikke innhenter sensitive personopplysninger.

Meld fra hvis du gjør vesentlige endringer i prosjektet

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

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

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

Vi tar kontakt om status for behandling av personopplysninger ved prosjektslutt

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Ved prosjektslutt 31.08.2018 vil vi ta kontakt for å avklare status for behandlingen av personopplysninger.

Se våre nettsider eller ta kontakt dersom du har spørsmål. Vi ønsker lykke til med prosjektet!

Marianne Høgetveit Myhren

Sri Tenden Myklebust

Kontaktperson: Sri Tenden Myklebust tlf: 55 58 22 68 / Sri.Myklebust@nsd.no

Vedlegg: Prosjektvurdering

Kopi: Susanne Dyrstad, susanndy@stud.ntnu.no

Vedlegg E: Samtykkeskriv

Susanne Dyrstad

<adresse>

<epost>

Til biologielever ved <navn på skole>

Anmodning om tillatelse til lydopptak av intervju.

Jeg er student på lektorprogrammet i realfag ved NTNU. Jeg ønsker å undersøke vurdering av skriftlige prøver i biologi.

For å få så godt dokumenterte data som mulig, er det ønskelig å gjøre lydopptak av intervjuet med deg og dine medstudenter. Derfor ber jeg om tillatelse fra deg til å kunne gjøre lydopptak. Det er snakk om ett intervju. Forutsetningen for tillatelsen er at alt innsamlet materiale blir behandlet med respekt og blir anonymisert, og at prosjektet ellers følger gjeldende retningslinjer for personvern. Det er helt frivillig å delta og man kan til enhver tid trekke seg fra deltakelse uten å måtte oppgi noen grunn til det.

Intervjuet vil bli utført i mars 2018, og intervjuet vil ha et omfang på omtrent tre kvarter. Opptakene vil kun bli hørt av meg og min veileder. I materiale som skrives eller på annen måte presenteres for andre vil involverte personer bli anonymisert. Innsamlede data vil bli slettet etter at prosjektet er avsluttet, senest 31. august 2018.

Hvis du vil vite mer om dette, eller hva det innsamlede materialet skal brukes til, så er det bare å ta kontakt med meg.

Jeg håper du synes dette er interessant og viktig, og at du er villig til å delta. Jeg ber deg om å fylle ut svarslippen på neste side hvis du sier ja til å bidra i prosjektet.

På forhånd takk!

Vennlig hilsen

Susanne Dyrstad

Tillatelse

Som del av masteroppgave ber jeg om tillatelse til å intervju deg sammen med fire andre elever hvor det blir gjort lydopptak av intervjuet.

Forutsetningen for tillatelsen er at innsamlet materiale blir anonymisert og behandlet med respekt, og at prosjektet følger gjeldende retningslinjer for etikk og personvern.

Sett kryss i ruta:

Jeg gir tillatelse.

Dato:

Ditt fornavn og etternavn: