

# **MASTEROPPGAVE**

**Masterstudium i skolerettet utdanningsvitenskap med  
fordypning i naturfagdidaktikk**

**Mai 2021**

*Kan kritisk tenkning i naturfag måles gjennom en skriftlig  
test?*

Sebastian Johansen



**OsloMet – storbyuniversitetet**

Fakultet for lærerutdanning og internasjonale studier

Institutt for grunnskole- og faglærerutdanning



## Sammendrag

Å kunne tenke kritisk er viktig for å kunne manøvrere seg gjennom dagens informasjonssamfunn, ta velbegrunnede avgjørelser for seg selv og samfunnet og for å mestre naturfaget. For å få innsikt i elevers evne til å tenke kritisk i naturfag, må dette kunne observeres. Studiens fokus er rettet mot de kognitive ferdighetene som er nødvendige for å kunne tenke kritisk i naturfag. Hensikten med dette masterprosjektet var derfor å utvikle og teste ut et kartleggingsverktøy for å måle elevenes ferdigheter innen kritisk tenkning i naturfag for ungdomstrinnet. Ferdighetstesten i kartleggingsverktøyet ble videreutviklet gjennom en suksessiv testutvikling, med fokus på autentiske tekster og problemstillinger. Masterprosjektet består derfor av fire studier hvor ferdighetstesten ble utprøvd, og hvor resultatene fra hver studie ble brukt for å revidere og kvalitetssikre kartleggingsverktøyet.

Resultatene fra det fjerde og siste studie indikerer at testscore ga et reliabelt mål på elevenes ferdighetsnivå, med utgangspunkt i de 72 elevene som deltok i studie 4. Det har ikke blitt gjennomført systematiske validitetsundersøkelser på testen. I likhet med eksisterende tester av kritisk tenkning gir ikke ferdighetstesten reliable resultater om de enkelte ferdighetene som måles, men kun som en sammenlagt testscore. Det er behov for videre studier og revideringer på ferdighetstesten dersom den skal kunne anvendes som et mer presist mål på de ferdighetene som er nødvendige for å tenke kritisk i naturfag, og for å få bedre oversikt over testens validitet og reliabilitet. Testen bør likevel anses som anvendelig ved masterprosjektets ende som et egnet supplement til naturfagslæreres inntrykk av elevenes evne til å tenke kritisk. Resultatene fra dette masterprosjektet kan ikke generaliseres, men bør anses som indikasjoner som bør undersøkes nærmere i et tilfeldig utvalg.



## Forord

Det er en sjelden mulighet å få bruke ett år av livet sitt på å fordype seg i et prosjekt, slik man gjør når man skriver en masteroppgave. Å lande dette prosjektet ville ikke vært mulig uten hjelp fra andre. Spesielt vil jeg takke min hovedveileder Kirsti Marie Jegstad. Jeg kunne ikke bedt om en bedre mentor. Hun har vist stor interesse for prosjektet, og brukt mye tid på å løse meg gjennom ukjent farvann mot land. Jeg har satt spesielt stor pris på hennes evne til å gi ærlige, gode og konstruktive tilbakemeldinger.

Det er et privilegium som masterstudent å ha tilgang på to veiledere, og jeg vil takke Marc Esser-Noethilichs for at du har bidratt med din erfaring og kunnskap om kvantitative forskningsmetoder. Resultatene fra dette prosjekter står stødigere takket være din kompetanse og støtte.

Uten lærere og elever som var villige til å delta i studiene ville ikke dette prosjektet vært mulig. Jeg vil derfor takke alle som har bidratt som respondenter til de fire studiene i prosjektet. I tillegg vil jeg takke tidligere elever som inspirerte meg til å starte på prosjekter gjennom å uttrykke stor interesse for kritisk tenkning.

Samfunnet har på mange måter stoppet opp det siste året, på grunn pandemi og smittevernstiltak. Dette har for min del muliggjort min helhjertede dedikasjon til dette prosjektet. Likevel vil jeg takke venner og familie som har vært tålmodige med meg gjennom året, og vist forståelse når det ikke har virket som det er nok timer i døgnet eller nok dager i studieåret. Jeg vil også takke alle som har bidratt med innspill og ved å lese korrektur på oppgaven og kartleggingsverktøyet. Til slutt vil jeg takke min bestemor som gikk bort denne våren, som alltid har støttet meg, og bidratt til å forme meg til den jeg er i dag.

Oslo

16.mai.2021

Sebastian Johansen



# Innholdsfortegnelse

<b>1</b>	<b>Innledning</b> .....	1
1.1	Bakgrunn for studien .....	1
1.2	Hensikt og forskningsspørsmål .....	4
1.3	Beskrivelse av kartleggingsverktøyet .....	5
<b>2</b>	<b>Teori</b> .....	7
2.1	Hva kjennetegner kritisk tenkning? .....	7
2.1.1	Forholdet mellom kritisk tenkning og fagkunnskaper .....	12
2.1.2	Kritisk lesning og argumentasjon .....	14
2.1.3	Hvorfor tenker vi ikke alltid kritisk? .....	17
2.2	Relevante oppgavetyper for kartleggingsverktøyet .....	19
2.2.1	Lukkede oppgaver .....	19
2.2.2	Åpne oppgaver .....	22
2.3	Klassisk testteori .....	23
2.3.1	Kriteriebasert test .....	23
2.3.2	Testens vanskelighetsgrad .....	24
2.3.3	Testens diskrimineringssevne .....	26
2.3.4	Reliabilitet i testutvikling .....	27
2.3.5	Validitet i testutvikling .....	29
2.3.6	Valg av oppgavetype med utgangspunkt i reliabilitet og validitet .....	30
2.4	Studier på eksisterende tester .....	32
<b>3</b>	<b>Metode</b> .....	36
3.1	Operasjonalisering og utvikling av kartleggingsverktøyet .....	37
3.1.1	Spørreskjema om bakgrunnsinformasjon .....	37
3.1.2	Test av ferdigheter .....	38
3.1.3	Spørreskjema om holdninger .....	44
3.2	Studiets oppbygging og utvalg .....	48
3.2.1	Studie 1 .....	50
3.2.2	Studie 2 .....	51
3.2.3	Studie 3 .....	52
3.2.4	Studie 4 .....	53
3.3	Analyse .....	54
3.3.1	Bearbeiding og klargjøring av datamaterialet .....	54
3.3.2	Kvalitetssikring av testen .....	55
3.3.3	Analyser av testens resultater .....	57

3.4	Etiske hensyn.....	59
<b>4</b>	<b>Resultater.....</b>	<b>60</b>
4.1	Studie 1.....	60
4.1.1	Del 1: Argumentsoppgaver.....	61
4.1.2	Del 2: Konklusjonsoppgaver .....	63
4.1.3	Del 3: Oppgaver om kildekritikk .....	64
4.2	Studie 2.....	65
4.3	Studie 3.....	67
4.3.1	Ferdighetstesten.....	68
4.3.2	Holdningskategoriene .....	72
4.4	Studie 4.....	73
4.4.1	Utvalg, frafall og dataregngjøring.....	73
4.4.2	Resultater knyttet til forskningsspørsmål 1: Kvalitetssikring av kartleggingsverktøyet	74
4.4.3	Resultater knyttet til Forskningsspørsmål 2: Testens resultater.....	87
<b>5</b>	<b>Diskusjon.....</b>	<b>98</b>
5.1	Hvor godt mål på elevers ferdigheter til å kritiske tenkning er ferdighetstesten? .....	98
5.1.1	Reliabilitet.....	98
5.1.2	Validitet .....	104
5.2	Testens resultater.....	109
5.3	Kritisk blikk på masterprosjektets metode.....	115
<b>6</b>	<b>Konklusjon .....</b>	<b>118</b>
6.1	Videre forskning .....	120
6.2	Implikasjoner for undervisning.....	121
	<b>Litteraturliste .....</b>	<b>122</b>
	<b>Vedlegg.....</b>	<b>129</b>



# 1 Innledning

## 1.1 Bakgrunn for studien

Kritisk tenkning har alltid vært viktig, men i dag lever vi i en spesiell tid i verdenshistorien. På grunn av teknologiske sprang har vi større tilgang på informasjon enn tidligere generasjoner. I løpet av noen dager produseres mer digital informasjon enn det man regner med at hele menneskeheten skrev frem til 2003 (Hansen, 2016). Mye av den informasjonen har vi tilgang til fra smarttelefoner, som får plass i lomma. Denne tilgangen gjør at informasjon kan utveksles raskt og enkelt. Dette byr på nye muligheter, men også utfordringer. Mens elever tidligere måtte lete gjennom bibliotekets bokhyller i sin søken etter informasjon, trenger dagens unge bare å gjøre noen få tastetrykk. Utfordringen er at med det tastetrykket vil elevene møte en overflod av informasjon. De må velge hvilken informasjon de skal forholde seg til, og disse valgene har vist seg å ofte være ukritiske og tankeløse. Elevene blir dermed passive mottakere (Frønes, 2017). Dette skjer samtidig som våre medievaner er i endring. Det er en økende trend i samfunnet er at det blir vanligere å motta dagens nyheter gjennom sosiale medier (Sakariassen et al., 2017). Dette kan føre til selektiv lesning, som øker behovet for å kunne lese kritisk. I dette informasjonssamfunnet har vårt forhold til kunnskap forandret seg. Før, da tilgangen på kunnskap var mindre tilgjengelig, kan man si at den kanskje hadde mer autoritet. Nå er det lett å oppleve strømmen av informasjon som kaotisk, og kunnskap som motstridende. Noen vil gå så langt som å påstå at vi lever i et post-faktuelt samfunn, hvor man har mistet troen på sannhet (Eriksen, 2018). Vi står i fare for at meninger og følelser får autoritet på bekostning av troen på empirisk vitenskap (Ferrer & Wetlesen, 2019).

Å utvikle elevenes kritiske tenkning er et viktig dannelsesoppdrag for skolen. Dette inkluderer evnen til kritisk lesning. Imidlertid finnes det indikasjoner på at norsk grunnskole ikke har mestret dette oppdraget. I midten av oktober 2020 kunne VG og aftenposten sjokkere leserne sine med overskrifter om at «*Mange norske elever sliter med kildekritikk*» (NTB, 2020) og «*Å være kritisk er viktigere enn noen gang. Men norske elever lærer ikke nok om det.*» (Stensland, 2020). Overskriftene kom i kjølvannet av en studie på PISA-resultatene til norske ungdomsskoleelever. Der fant Weyergang og Frønes (2020) at norske elever mangler strategier for å avgjøre om de kan stole på en kilde eller ikke.

En del av informasjonen som elevene møter vil ha et naturvitenskapelig aspekt. Ettersom samfunnet endrer seg stadig raskere har det, ifølge Millar og Osborne (1998), blitt vanskeligere å forutse hva slags kunnskaper elevene trenger fra naturfaget. Det vi kan forutse er at elevene i fremtiden vil møte problemstillinger med naturvitenskapelige aspekter. I media kan man daglig lese om sosiovitenskapelige kontroverser der det ikke finnes ett fasitsvar, som for eksempel global oppvarming, genteknologi, rusproblematikk, ulvedebatten eller utfordringer med livsstilsykdommer. Det mest nærliggende eksemplet på et sosiovitenskapelig kontrovers er hvordan det siste året har vært preget av å leve under en pandemi, med utstrakt debatt om vaksiner, smittevernstiltak og hvordan samfunnet bør prioritere. I møte med slike problemstillinger er det viktig at elevene tar velbegrunnede avgjørelser for både samfunnets del og for seg selv. Hvis elevene forblir passive informasjonsmottakere kan de fort begynne å tro på påstander med lite empirisk bakgrunn, og i verste fall en konspirasjonsteori, og hvis man først har akseptert en vrangforestilling, kan det være lettere å falle for en annen (Hope & Klausen, 2012). Dette gjør naturfagets allmenndannende aspekt enda viktigere, som betyr at faget bør gjøres relevant også for de som ikke søker et yrke innen naturvitenskapen (Sjøberg, 2009).

Spørsmålet er om dagens naturfagundervisning egentlig forbereder elevene på å kunne orientere seg i informasjonssamfunnet og forholde seg til sosiovitenskapelige kontroverser. Kolstø (2006) trekker frem at kompetanse innen kritisk tenkning i naturfag er viktigere enn tidligere, fordi det har skjedd endringer ved naturvitenskapen. Naturvitenskapelig informasjon har endret seg mot å legitimere og begrunne beslutninger innen industri eller samfunnsliv. Før var forskere ofte ansatt ved uavhengige universiteter, mens de fleste forskere og naturvitere nå er ansatt i private selskaper, industrien eller tilknyttet andre offentlige institusjoner, og Kolstø (2006) kaller dette for "*industrialisert naturvitenskap*". Det er naturlig å stille seg kritisk til om det kan oppstå interessekonflikter i den industrialiserte naturvitenskapen. Utenom skolen vil elevene stort sett møte på vitenskapelige påstander produsert gjennom industrialisert naturvitenskap, selv om skolen i hovedsak har fokusert på universitetsbasert akademisk vitenskap (Kolstø, 2006). Det er altså et økt behov for å øve opp elevene i kritisk tenkning rundt naturvitenskapelig informasjon.

Disse tankene er i tråd med det Vieira og Tenreiro-Vieira (2016) anser som selve målet med naturfag; *scientific literacy*. Selv om det ikke er konsensus om hva som konkret menes med *scientific literacy*, knytter flere definisjoner begrepet til å kunne bruke kunnskap om

vitenskapelige prosesser og ideer til å forstå naturvitenskapelig informasjon og å ta gode avgjørelser basert på slik informasjon. For å oppnå scientific literacy er kritisk tenkning grunnleggende (Vieira et al., 2011). For å oppsummere er det to hovedgrunner til at kritisk tenkning er viktig i naturfag. For det første er kritisk tenkning viktig for å mestre faget, fordi kritisk tenkning ligger i kjernen av vitenskapelige tenkemåter og scientific literacy. For det andre er kritisk tenkning viktig for å kunne ta velbegrunnede avgjørelser i en verden hvor vi bombarderes med naturvitenskapelige påstander.

Naturfaglærere står dermed med et stort ansvar, ettersom kritisk tenkning kan anses som den mest sentrale ideen for hele opplæringen i et liberalt demokrati (Vieira et al., 2011). Behovet for innsikt i vitenskapelige tenkemåter og kritisk tenkning adresseres i overordnet del av fagfornyelsen som ble innført høsten 2020. «*Kritisk tenkning og etisk bevissthet*» er et av seks punkter som skal danne verdigrunnlaget for hele opplæringen (Utdanningsdirektoratet, 2019b). Skolen skal bidra til at elevene utvikler «*vitenskapelig og kritisk tenkning*», og «*elevene skal kunne vurdere ulike kilder til kunnskap og tenke kritisk om hvordan kunnskap utvikles. De skal også kunne forstå at deres egne erfaringer, standpunkter og overbevisninger kan være ufullstendige eller feilaktige. Kritisk refleksjon forutsetter kunnskap, men gir samtidig rom for usikkerhet og uforutsigbarhet. Opplæringen må derfor søke en balanse mellom respekt for etablert viten og den utforskende og kreative tenkningen som kreves for å utvikle ny kunnskap.*» (Utdanningsdirektoratet, 2019c). Slik jeg tolker dette må naturfag spille en særrolle i dette krevende samfunnsoppdraget, for naturvitenskapen har lange historiske røtter og er tett knyttet opp mot vitenskapsfilosofi. Dessuten vil elevene måtte forholde seg til mange naturvitenskapelige påstander, og ta veloverveide avgjørelser basert på naturvitenskapelig informasjon gjennom livet.

Et relevant spørsmål å stille seg er hvordan lærere, elever og myndigheter kan vite om de har utviklet elevens evne til å tenke kritisk. Etter tusenårsskifte har nasjonale og internasjonale standardiserte tester fått en større rolle i skolen og i skoledebatten (Dahl et al., 2016).

Hensikten med slike tester er å bidra til å øke innsikten i elevenes kompetanse, og hvordan slik kompetanse kan utvikles på best mulig måte (Schleicher, 2019; Utdanningsdirektoratet, u.å.). Spørsmålet blir i hvilken grad tester kan fange helheten i elevenes kompetanse. Et sitat fra 1600-talls fysikeren og astronomen Galileo Galilei fanger kanskje tidsånden i skolen:

«*...measure what is measurable and make measurable what cannot be measured*»  
(Finkelstein, 1982, s. 1).

Som lærer har jeg vært interessert i temaet kritisk tenkning, både fordi det er svært relevant for elevenes hverdag og fremtid, men også fordi det er et konsept som kan fremstå tvetydig og uklart. Min intensjon er derfor å undersøke om elevenes kritiske tenkning kan observeres, slik at det er mulig å få innblikk i hvordan man kan bedrive god opplæring i kritisk tenkning innenfor naturfaget. Måling av kritisk tenkning har vært undersøkt før, men det finnes få tilgjengelige tester som man kan anvende som lærer, og de kommersielle testene som måler kritisk tenkning fokuserer ikke på enkeltfag, og er lite tilgjengelige grunnet opphavsrett og man må betale for å få tilgang til dem.

## 1.2 Hensikt og forskningsspørsmål

Hensikten med prosjektet er å utvikle og teste ut et kartleggingsverktøy for å måle elevenes ferdigheter innen kritisk tenkning i naturfag for ungdomstrinnet. Masteroppgavens hovedfokus vil være rettet mot selve testutviklingen. Kartleggingsverktøyet skal gi innsikt i elevenes kritiske tenkning om naturfaglig informasjon om autentiske problemstillinger, og hva som påvirker slik tenkning, og resultatene fra utprøvingen vil bli analysert for å få innsikt i dette.

For å undersøke dette vil masteroppgaven ta utgangspunkt i to forskningsspørsmål (FS):

1. Hvordan kan man teste ferdighetene som er nødvendige for å tenke kritisk i naturfag på ungdomstrinnet?
2. Hva kjennetegner testens resultater?
  - a. Hvordan er vanskelighetsgraden til de ulike delferdighetene og testen som helhet?
  - b. Hvordan skiller testen mellom ulike elevgrupper (kjønn, trinn, minoritetspråklig bakgrunn, karakter i naturfag og norsk og holdninger til kritisk tenkning)?

Et utgangspunkt for kartleggingsverktøyet er at kritisk tenkning kan ses på som en kompetanse, som består av kognitive ferdigheter og holdninger (Facione, 1990a). Etter en slik forståelse av kritisk tenkning vil ferdighetene være det elevene anvender for å tenke kritisk, mens holdningene til kritisk tenkning vil være førende for hvilke situasjoner elevene vil tenke kritisk. Kapittel 2 gir en mer detaljert redegjørelse for det teoretiske rammeverket for studien. Prosjektet kan potensielt gi naturfagslærere mulighet til å kartlegge elevenes kritiske

tenkning, og få innsikt i hva de bør fokusere på i undervisningen for å styrke elevenes ferdigheter til å tenke kritisk i naturfag. Hovedfokuset i dette masterprosjektet vil være på elevenes ferdigheter, som altså kun er en del av kompetansen som er viktig for å tenke kritisk. Hvis lærere skal finne effektive undervisningsmetoder og strategier for å utvikle elevenes kompetanse til kritisk tenkning vil det være av stor betydning om ferdighetene kan måles.

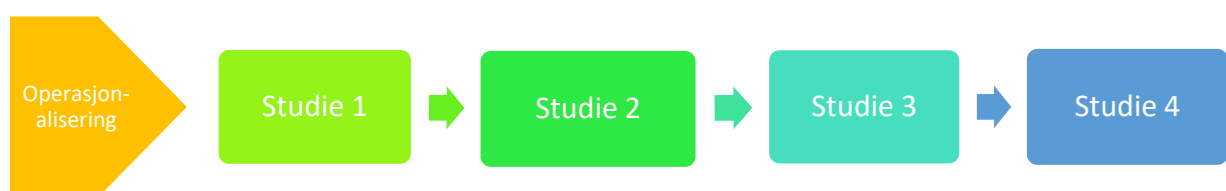
### 1.3 Beskrivelse av kartleggingsverktøyet

Kartleggingsverktøyet består av tre deler, som vises i Figur 1-1. Den første delen av kartleggingsverktøyet består av et spørreskjema som undersøker variabler som kjønn, trinn, etnisitet og terminkarakter i naturfag og norsk. Variablene ble brukt for å undersøke mulige korrelasjoner med elevenes testscore. Den andre delen av kartleggingsverktøyet har som hensikt å teste elevenes ferdigheter til å tenke kritisk i naturfag. Utviklingen og testingen av denne delen av kartleggingsverktøyet er hensikten med masterprosjektet. Derfor vil ferdighetstesten tilegnes mest fokus gjennom oppgaven. Ett av målene med ferdighetstesten i kartleggingsverktøyet er å basere oppgavene på virkelighetsnære problemstillinger, med fokus på kildekritikk av autentiske tekster og å kunne fatte velbegrunnede konklusjoner. Kartleggingsverktøyets tredje del består av et spørreskjema som undersøker elevenes holdninger til kritisk tenkning, og naturfag. Dette gjøres for å kunne sammenligne elevenes score i ferdighetstesten og elevenes holdninger. Dette kan gi et mer sammensatt bilde av elevenes kompetanse innen kritisk tenkning i naturfag, men siden hovedfokuset i dette masterprosjektet er ferdigheter vil det gjennomføres mest kvalitetsundersøkelser på ferdighetstesten.



Figur 1-1: Kartleggingsverktøyets tre deler

Utviklingen av kartleggingsverktøyet vil gjøres gjennom en *suksessiv testutvikling*, som vist i Figur 1-2. Det betyr at masterprosjektet består av flere utprøvnings av måleverktøyet, hvor resultatene fra hver studie blir brukt for å revidere og kvalitetssikre kartleggingsverktøyet. Både utviklingsprosessen og selve kartleggingsverktøyet er ytterligere redegjort for i kapittel 3.



Figur 1-2: Prosjektets fire stadier for testutvikling

## 2 Teori

I dette kapittelet vil jeg gjøre rede for det teoretiske fundamentet i utviklingen av kartleggingsverktøyet for kritisk tenkning i naturfag. Først beskrives og drøftes ulike synspunkt på hva som kjennetegner kritisk tenkning. Deretter redegjøres det for hvorfor kildekritikk og argumentasjon er sentrale elementer innen kritisk tenkning. Mot slutten av kapittelet tar jeg for meg anbefalinger om hvordan man kan utvikle en test og undersøke testens validitet og reliabilitet. I tillegg forsøker jeg å gjøre rede for hva studier på eksisterende tester om kritisk tenkning har funnet ut.

### 2.1 Hva kjennetegner kritisk tenkning?

Det finnes ingen enighet blant akademikere om hva som ligger i begrepet kritisk tenkning (Bailin, 2002). Det kan virke påfallende, ettersom begrepet har tradisjoner helt tilbake til de antikke tenkerne. Til tross for at kritisk tenkning kan ha mange betydninger blir det mye brukt i ulike utdanningspolitiske dokumenter (Brodin, 2007). Dette fører til at det sjeldent er en tydelig rød tråd i hvordan begrepet brukes i slike publikasjoner (Brodin, 2007). Selv om begrepet kritisk tenkning har fått større plass i fagfornyelsen enn i tidligere læreplaner, kan det diskuteres i hvor stor grad det er konkret definert, foruten at det har noe med kildekritikk, granskning av ideer, kunnskap og kritisk refleksjon av hvordan ny kunnskap blir til å gjøre (Utdanningsdirektoratet, 2019c).

Det mangler ikke meninger, akademisk diskusjon eller forskningslitteratur på emnet. De som ikke er akademikere forbinder gjerne begrepet med det å kritisere, altså gi en negativt ladet kritikk (Eriksson & Blomgren, 2020). I den akademiske diskusjonen er det vanlig å se på kritisk tenkning som et mer omfattende konsept enn dette. Noen har forsøkt å konkretisere kritisk tenkning som evnen og viljen til å kunne utføre en rekke kognitive prosedyrer. Dette perspektivet kritiserer Bailin (2002) for å være «feiloppfatninger» om hva kritisk tenkning er: «*This approach characterizes critical thinking in descriptive terms, describing a range of behaviours or activities, for example the steps involved in problemsolving or the stages of inquiry. Yet simply carrying out a set of procedures is not sufficient to ensure critical thinking since any procedure can be carried out carelessly, superficially, or unreflectively – in other words, in an uncritical manner.*» (Bailin, 2002, s. 363). Bailin mener at det er kvaliteten på tenkning som er viktig, og ikke hvilke utvalgte prosedyrer som blir gjennomført. For henne blir et slikt syn på kritisk tenkning for snevert. Et viktig poeng er at kritisk tenkning ikke kan

reduseres til en kokebokoppskrift for hvilke prosedyrer som må gjennomføres (Bailin, 2002). Hun argumenterer videre for at en god definisjon av kritisk tenkning bør innebære et normativt kriterium for kritisk tenkning: «*A non-normative description of a procedure fails to capture what is most essential about critical thinking*» (Bailin, 2002, s. 364). Slik som det ble beskrevet i masteroppgavens innledning er det viktig at elevene kan orientere seg i et samfunn hvor de blir bombardert med naturfagsrelaterte påstander, og hvor de må ta avgjørelser om sosiovitenskapelige kontroverser, som påvirker både samfunnet og dem selv. Flere i den akademiske diskusjonen har pekt på at det normative målet med kritisk tenkning er at slike avgjørelser skal være så velbegrunnede som mulig.

Et slikt perspektiv kan vi spore tilbake til utdanningsfilosofen Dewey. I boken «*How we think*» skiller han mellom to typer tenkning. Den ene formen for tenkning er ukritisk. Slik tenkning kjennetegnes ved at man passivt aksepterer ideer og informasjon som blir presentert, og man trekker forhastede konklusjoner uten å granske bevis eller reflektere over og vurdere grunnlaget for konklusjonen. Reflekterende tenkning kjennetegnes derimot av en «*Active, persistent and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it, and the further conclusions to which it tends*» (Dewey, 1910, s. 6). Reflekterende tenkning kan altså oppsummeres ved at man aktivt reflekterer over påstander og ideer, blant annet gjennom å vurdere beleggene for påstanden og reflekterer over hvilke logiske konsekvenser påstanden har. For Dewey er målet med slik tenkning å ta logiske avgjørelser ut fra de bevisene, kunnskapene og verdiene man har tilgjengelig. Definisjonen av reflekterende tenkning kan minne om Ennis sin anerkjente definisjon av kritisk tenkning: «*Critical thinking is reflective and reasonable thinking that is focused on deciding what to believe or do*» (Ennis, 1985, s. 45). Denne definisjon har et normativt mål for kritisk tenkning, hvor målet er at den som tenker skal være egnet til å ta en veloverveid avgjørelse om hva han skal gjøre eller tro. Det er slike normative mål Bailin savner i en del andre definisjoner.

Kritisk tenkning består av to elementer, ifølge Ennis (1985); kognitive ferdigheter (abilities) og holdninger (dispositions). Ideen om at kritisk tenkning består av ferdigheter og holdninger ble styrket gjennom et omfattende delphistudie, ved navn «The Delphi Report» som ble gjennomført av Facione (1990a). Studien hadde som hensikt å definere hva kritisk tenkning er, og hva som kjennetegner det en gjennomsnittlig collegestudent bør evne av kritisk tenkning. Delphistudien gjennomgikk flere runder med forslag om hva som ligger i begrepet



kritisk tenkning, slik at det kunne utvikles en konsensus blant ekspertene. De 46 ekspertene som deltok i studien hadde bakgrunn i filosofi, pedagogikk og andre relevante fagfelt. De forsøkte å formulere og bli enige om kjennetegn på kritisk tenkning (Facione, 2015). Panelet med eksperter ble enige om følgende beskrivelse av kritisk tenkning:

*«We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgement is based.»* (Facione, 1990a, s. 2)

Hensikten med å beskrive kompetansen kritisk tenkning som ferdigheter og holdninger er at begrepet er enklere å konkretisere, analysere og måle (Facione, 1990a). Et individ må inneha både ferdighetene og holdningene for å kunne bestemme seg for hva de skal tro og gjøre, altså målet med kritisk tenkning. Ferdighetene kan ses på som de kognitive verktøyene som er nødvendige for å kunne utføre kritisk tenkning. Holdningene kan forklares som innstillinger individet har til ny informasjon, enten generelt eller temaspesifikt. Individets holdninger avgjør dermed om ferdighetene til kritisk tenkning anvendes (Facione, 1990a). Eksempler på slike holdninger er ønsket om å være velinformert, ha tillit til grundige og systematiske undersøkelser, et åpent sinn til ny informasjon og et ønske om å endre mening hvis andre alternativer er mer sannsynlige enn det standpunktet man har selv. Facione (2015) beskriver en person med slike holdninger som å ha «*a critical spirit*». Med utgangspunkt i en slik forståelse kan altså ikke en ferdighetstest i kritisk tenkning forutsi om elever vil opptre som kritisk tenkere utenom testsituasjonen, selv om de har evnen til å gjøre det.

I delphistudien ble det formulert en rekke eksempler på holdninger knyttet til kritisk tenkning (Facione, 1990a), og med utgangspunkt i ekspertpanelets arbeid har Facione et al. (2000) kategorisert holdningene inn i syv kategorier. Holdningskategoriene fremstilles i Tabell 2-1, og har blitt oversatt av meg.

Tabell 2-1: Holdningskategorier anvendt av (Facione et al., 2000, s. 2-3). Oversatt av Sebastian Johansen

Holdning	Beskrivelse
<b>Intellektuell nysgjerrighet</b> (Inquisitiveness)	Holdninger som omhandler personens intellektuelle nysgjerrighet.
<b>Åpenhet for ny informasjon</b> (Open-Mindedness)	Holdninger knyttet til personens åpenhet for ny informasjon, og andres meninger, selv om dette kan motstride egne verdier og meninger.
<b>Selvtillit til egen evne til å tenke kritisk</b> (Critical Thinking Self-Confidence)	Holdninger knyttet til tiltro til egne evner til å løse situasjoner hvor man må anvende kognitive ferdigheter til å tenke kritisk.
<b>Analytisk</b> (Analyticity)	Holdninger knyttet til om personen har tiltro til egen overvåkenhet ovenfor potensielle problemer, situasjoner, konsekvenser og muligheter.
<b>Sannhetssøkende</b> (Truth-Seeking)	Holdninger knyttet til personens ønske om å søke til sannheten, selv om dette kan medføre at respondenten selv må endre holdninger, verdier og meninger.
<b>Grundighet / systematisk</b> (Systematicity)	Holdninger knyttet til å undersøke problemstillinger systematisk og fokusert.
<b>Modenhet</b> (Maturity)	Holdninger knyttet til om en person tar reflekterte og gjennomtenkte avgjørelser og bedømmelser.

Delphistudien peker også på seks kognitive ferdigheter og seksten delferdigheter som de knytter til kritisk tenkning. Siden ferdighetene er sentrale i mitt masterprosjekt, har jeg oppsummert og oversatt de slik de presenteres i Tabell 2-2. I oppsummeringen har jeg forsøkt å velge ut det viktigste meningsinnholdet og forkortet beskrivelsen av ferdighetene og delferdighetene. For eksempel anvender Facione (1990a) denne beskrivelsen av ferdigheten “Tolke”: «*To comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgements, conventions, beliefs, rules, procedures or criteria*» (Facione, 1990a, s. 13). Jeg har valgt å oppsummere dette som: «*Tolke: Å kunne forstå og vise forståelse for en rekke ulike former for informasjon.*».

Tabell 2-2: Ferdigheter og delferdighetene som ifølge Facione (1990a) er sentrale for kritisk tenkning. Oversatt og oppsummert av Sebastian Johansen

Ferdighet	Delferdigheter
<p><b>Tolke (Interpretation)</b> Å kunne forstå og vise forståelse for en rekke ulike former for informasjon.</p>	<p><b>a. Kategorisere</b></p> <ul style="list-style-type: none"> <li>• Lage kategorier ut fra informasjonen og på den måten skille mellom ulike typer informasjon</li> </ul> <p><b>b. Avkode</b></p> <ul style="list-style-type: none"> <li>• Forstå informasjonen i den konteksten den uttrykkes, med tanke på for eksempel hvilke interesser den som uttrykker informasjonen har og hvordan den uttrykkes.</li> </ul> <p><b>c. Tydeliggjøre mening</b></p> <ul style="list-style-type: none"> <li>• Omformulere innholdet fra en informasjonskilde for å tydeliggjøre budskapet.</li> </ul>
<p><b>Analysere (analysis)</b> Å kunne identifisere og analysere innholdet i argumentasjon.</p>	<p><b>a. Undersøke ideer</b></p> <ul style="list-style-type: none"> <li>• Fastslå hvilken rolle ulike deler i et argument spiller eller er ment til å spille.</li> </ul> <p><b>b. Identifisere argumenter</b></p> <ul style="list-style-type: none"> <li>• Identifisere hvilken konklusjon et argument er ment for å støtte.</li> </ul> <p><b>c. Analysere argumenter</b></p> <ul style="list-style-type: none"> <li>• Kan beskrive hvordan en konklusjon begrunnes gjennom å analysere de ulike elementene i argumentet.</li> </ul>
<p><b>Vurdere (evaluating)</b> Å kunne vurdere troverdigheten og den logiske styrken i et argument eller en påstand.</p>	<p><b>a. Vurdering av troverdighet</b></p> <ul style="list-style-type: none"> <li>• Identifisere faktorer som er relevante for å vurdere troverdigheten til en kilde eller en påstand.</li> </ul> <p><b>b. Vurdering av argumenter</b></p> <ul style="list-style-type: none"> <li>• Vurdere styrken og relevansen til premissene for en konklusjon.</li> <li>• Reflektere over logiske konsekvenser av premissene og påstandene i argumentasjonen.</li> </ul>
<p><b>Å trekke konklusjoner (inference)</b> Å kunne identifisere hva som trengs for å trekke troverdige konklusjoner, og bruke relevant informasjon for å trekke gjennomtenkte slutninger.</p>	<p><b>a. Undersøke bevis</b></p> <ul style="list-style-type: none"> <li>• Identifisere nødvendige elementer for å trekke en godt begrunnet konklusjon, og undersøke disse.</li> </ul> <p><b>b. Formulere alternativer</b></p> <ul style="list-style-type: none"> <li>• Formulere alternative hypoteser og forslag, og vurdere om disse kan være gode forklaringer.</li> </ul> <p><b>c. Trekke konklusjoner</b></p> <ul style="list-style-type: none"> <li>• Kunne bruke og sette sammen ulike typer bevis for å trekke en konklusjon som er logisk gyldig.</li> <li>• Kunne vurdere flere mulige konklusjoner opp mot hverandre.</li> </ul>
<p><b>Forklare (explanation)</b> Å kunne forklare og rettfærdiggjøre hvordan en har kommet frem til en konklusjon.</p>	<p><b>a. Beskrive resultater</b></p> <ul style="list-style-type: none"> <li>• Beskrive hvilke konklusjoner man har kommet frem til, og hvordan man kom frem til dem.</li> </ul> <p><b>b. Rettfærdiggjøre prosedyrer</b></p> <ul style="list-style-type: none"> <li>• Forklare hvilke valg man gjorde for å undersøke en problemstilling, og kunne begrunne hvorfor man tok disse valgene.</li> </ul> <p><b>c. Presentere argumenter</b></p> <ul style="list-style-type: none"> <li>• Gjøre rede for hvorfor man støtter en konklusjon.</li> <li>• Kunne forutse og svare på motargumenter av konklusjonen man har trukket.</li> </ul>

<p><b>Selvregulering (Self-regulation)</b>  Å kunne være bevisst på hvordan en selv tenker, og stille seg kritisk til egne valg, viten og holdninger.</p>	<p><b>a. Selvrefleksivitet</b></p> <ul style="list-style-type: none"> <li>• Reflektere over hvordan man selv kommer frem til en konklusjon eller mening, og vurderer kritisk hva som kan påvirke egne meninger.</li> </ul> <p><b>b. Selvkorrigerings</b></p> <ul style="list-style-type: none"> <li>• Er innstilt på å korrigere for mangler hos seg selv og egne resonnmener.</li> <li>• Utvikler strategier for å rette opp i disse.</li> </ul>
---	---

Fordelen med å dele begrepet kritisk tenkning inn i ferdigheter og holdninger er at begrepet blir konkret, samtidig som det som det kan romme mange elementer. Faciones (1990a) tolkning av kritisk tenkning egner seg på denne måten bedre til å operasjonalisere i et måleinstrument enn bredere definisjoner. Derfor har jeg valgt å ta utgangspunkt i holdningskategoriene og ferdighetskategoriene nevnt over for å utvikle kartleggingsverktøy som skal måle kritisk tenkning innen naturfag.

### 2.1.1 Forholdet mellom kritisk tenkning og fagkunnskaper

I hvor stor grad kritisk tenkning er knyttet til fagkunnskap er et omdiskutert spørsmål. Den lengste avstanden finner vi mellom de som mener at kritisk tenkning er generelle tenkeferdigheter, og de som mener at kritisk tenkning ikke kan skilles fra faglig innhold. Fisher (2011) knytter kritisk tenkning til kun kognitive ferdigheter som er grunnleggende for all kritisk tenkning, og representerer den første posisjonen. Han argumenterer for at kritisk tenkning bør anses som en egen disiplin, separert fra fag. McPeck (1985) inntar den motsatte posisjonen av Fisher, ettersom han argumenterer for at kritisk tenkning ikke kan ses separat fra faglig innhold. Å forsøke å beskrive kritisk tenkning, slik den anvendes i ulike fagdisipliner, som generelle ferdigheter basert på logikk, vil ifølge McPeck (1990) føre til at ferdighetene blir for generelle for å anvendes til ulikt faglig innhold.

Mellom de to posisjonene, finnes en rekke mellomposisjoner. Ennis (1987) argumenterer for at elevenes evne til kritisk tenkning sannsynligvis kan overføres mellom fag og temaer. Ifølge Ennis (1987) forventes det at vi klarer å overføre mye av det vi lærer i fagene på skolen til andre situasjoner senere i livet. Samtidig anerkjenner Ennis at lærere ikke bør forvente at elevene klarer å overføre de kognitive ferdighetene knyttet til kritisk tenkning mellom fag og temaer selv, og derfor er det viktig at lærere øver dem opp til det. Slik jeg tolker Facione (2015), virker det som at han har et lignende syn på kritisk tenkning. I en forelesning viser Facione (2018) til ett metastudie gjennomført av Abrami et al. (2008), som har undersøkt effekten av eksplisitt og implisitt undervisning i kritisk tenkning. Studien viser klare indikasjoner på at eksplisitt undervisning om kritisk tenkning, i form av egne kurs for

studenter ved college, gir større effekt på studentenes evne til å tenke kritisk enn når det undervises om kritisk tenkning gjennom fagene. Likevel anerkjenner Facione at fagkunnskaper er viktig for kritisk tenkning: «*Too much of value is lost if CT is conceived of simply as a list of logical operations and domain-specific knowledge is conceived of simply as an aggregation of information.*» (Facione, 1990a, s. 10). Bailin (2002) mener derimot at kritisk tenkning består av temaspesifikke ferdigheter, og at kritisk tenkning alltid bør ses i sammenheng med temaet det undervises i. Hun argumenterer for at en analyse av litterære tekster i språkfag krever andre ferdigheter enn det som kreves i naturfag (Bailin, 2002). Ifølge et slikt syn kan kritisk tenkning anses som et paraplybegrep som viser til kvaliteten på tenkningen.

Perspektivene over viser at det ikke er konsensus om hva som kjennetegner en kritisk tenker. Definisjonene til Ennis og Facione innebærer det normative målet som Bailin etterspør. Ferdighetene slik de er beskrevet av Facione (1990a) er mer enn prosesser som elevene må gjennomføre, og kjennetegnene på ferdighetene er formulert slik at man kan skille mellom kritisk tenkning av høy og lav kvalitet. Om kritisk tenkning er en generell eller temaspesifikk ferdighet kommer ikke denne masteroppgaven til å finne et endelig svar på, men jeg vil påstå at ferdighetene i Tabell 2-2 er relevante for kritisk tenkning i naturfag, og grunnleggende for å utvikle scientific literacy. Ferdighetene er tett knyttet opp mot det Hope og Klausen (2012) kaller vitenskapelig skepsis. Naturfag har en særrolle i oppdraget om å lære elevene om vitenskapelige tenkemåter, og kritisk tenkning er en forutsetning for vitenskapelig tenkning. Derfor vil ferdighetene, slik de formuleres i Facione (1990a) anvendes som et utgangspunkt for kartleggingsverktøyet for kritisk tenkning. Kriteriene for ferdighetene, slik de er formulert i Tabell 2-2, må tilpasses for elever på ungdomstrinnet, fordi de er ment for collegestudenter.

Vieira og Tenreiro-Vieira (2016) har forsøkt å definere kritisk tenkning i naturfag, og henter inspirasjon fra Facione (1990a), Ennis (1985) og Bailin (2002). Ifølge Vieira og Tenreiro-Vieira (2016) er kritisk tenkning avhengig av kognitive ferdigheter og holdninger, men også naturfaglige kriterier og fagkunnskaper, som vist i Figur 2-1. Å tenke kritisk i naturfag er når elevene forstår de naturfaglige aspektene ved en problemstilling, og evner å ta en veloverveid avgjørelse om problemstillingen gjennom å vurdere tilgjengelig informasjon ut fra kriterier som validitet, relevans og påliteligheten til informasjonskildene (Vieira et al., 2011). Relevant fagkunnskap kan være kjennskap til vitenskapelige metoder og tenkemåter, sentrale ideer innen naturvitenskap eller kunnskaper om spesifikke naturfaglige temaer (Vieira & Tenreiro-



naturvitenskaplig påstander som elevene møter på i ulike medier være motstridende. Fordi elevene i så stor grad vil måtte forholde seg til tekstbaserte kilder, er en sentral del av det å tenke kritisk knyttet opp mot det å kunne lese kritisk (Weyergang & Frønes, 2020). I fagfornyelsen presiseres det at det er et av skolens oppdrag å lære elevene å kunne vurdere ulike kilder og tenke kritisk om hvordan kunnskap utvikles (Utdanningsdirektoratet, 2019c). Det kan virke som at skolen ikke har løst dette oppdraget. I en studie av resultater fra PISA 2018 finner Weyergang og Frønes (2020) at norske elever mangler strategier for å vurdere påliteligheten i ulike tekster. De argumenterer for at når skolen ikke gir god nok opplæring i kildekritikk, overlates opplæringen til tilfeldighetene, og hjemmets tekstpraksiser får unødvendig stor innflytelse på elevenes evne til å lese kritisk. Selv om det er elever med generelt gode leseferdigheter som er flinke til å vurdere informasjonskilders troverdighet (Frønes, 2017), trekker Weyergang og Frønes (2020) frem at det er god grunn til å tro at alle kan lære å inneha en kritisk innstilling i møte med tekster. Kritisk lesning trenger bare å innebære at man stiller noen grunnleggende spørsmål til tekstene. Elevene kan komme langt ved å evne å plassere seg utenfor teksten og stille seg spørsmål om for eksempel hva som er hensikten med teksten, hvem er og hvilket motiv har forfatterne, og hvordan påstandene er begrunnet (Weyergang & Frønes, 2020). I en tekst fremlegges det som regel argumenter for å begrunne påstander. Det betyr at for å kunne vurdere om en påstand er gyldig eller ikke er må man undersøke hvordan det argumenteres i teksten, og innholdet i argumentene som fremlegges i teksten (Walton, 2005).

Spørsmålet blir hva som kjennetegner god argumentasjon i naturfag. Argumentasjon er et sentralt kjennetegn på naturvitenskapen, og driver vitenskapen frem mot kunnskapsproduksjon (Mork, 2008b). Mork (2008b) mener kompetanse innen argumentasjon er tett tilknyttet utviklingen av kritisk tenkning og andre avanserte kognitive ferdigheter. Facione (1990a) mener også at evnen til å kunne forstå og vurdere argumenter er viktig for kritisk tenkning, og argumenter er nevnt eksplisitt i enkelte av delferdighetene som presenteres i Tabell 2-2. Et argument kan kort oppsummeres som en påstand og beleggene som fremlegges som begrunnelser for påstanden (Mork, 2008a; Toulmin, 2003). Det kan legges flere elementer til argumentet som har til hensikt å styrke relevansen mellom påstanden og belegget (Toulmin, 2003).

For å kunne vurdere og formulere argumenter er det viktig å vite hva som kjennetegner et godt argument. Diskusjonen om hva som definerer et gyldig og sterkt argument har vært

diskutert siden før Sokrates tid. Et sterkt argument er logisk gyldig. Logikk handler om evnen til å trekke korrekte slutninger fra gitte premisser (Eriksson & Blomgren, 2020). Hvis et argument er logisk gyldig, vil beleggene og premissene som fremlegges bevise påstanden i argumentet. Derfor er vanlige kriterier for sterke argumenter at beleggene i argumentet har høy troverdighet, er logisk gyldige og er relevante for påstanden (Eriksson & Blomgren, 2020). Noen vanlige argumenter er logisk ugyldige, men mange lar seg overbevise av dem. Disse kalles *fallasier*, eller tankefeil (Walton, 2010). Ifølge Hansen (2002) kan fallasier defineres som argumenter som fremstår bedre enn de egentlig er. Vanlige fallasier kan være autoritetsargumenter og popularitetsargumenter, hvor man hevder at en påstand er sann, fordi den kommer fra en autoritetsperson eller det er mange som mener det samme. En annen type fallasier er skråplansargumenter, hvor man hevder at en avgjørelse i en bestemt retning vil føre til det verst tenkelige scenarioet i den retningen (Eriksson & Blomgren, 2020).

At fallasier er så utbredt, kan indikere at det er et stort behov for kunnskaper om god argumentasjon. Ettersom kompetanse i argumentasjon er knyttet til kildekritikk, er det viktig at elevene kan identifisere argumenter i tekster og vurdere hvor god argumentasjonen er. Dette kan gjøres gjennom å undersøke om begrunnelsene er relevante for påstanden og om de er logisk gyldige. Det er også viktig at premissene i argumentet er stemmer. Sannsynligheten for at premissene er riktige påvirkes av troverdigheten til kilden. Denne kan vurderes ved å undersøke om forfatteren faktisk er en ekspert på temaet, hvor troverdig han er som ekspert, hva slags bevis påstandene bygger på og om dette stemmer overens med andre eksperter på temaet. I tillegg kan det være relevant å stille spørsmål ved om informasjonen stemmer overens med kunnskapen en selv allerede har om temaet (Walton, 2005).

Kompetanse i kildekritikk og argumentasjon er viktig for kritisk tenkning generelt og i naturfag. Derfor har jeg valgt å fokusere på kildekritikk og vurdering av argumenter når kritisk tenkning i naturfag har blitt operasjonalisert i kartleggingsverktøyet. Kildekritikk er viktig, spesielt i forhold til nettsider, for å utvikle elevenes scientific literacy og evne til å ta veloverveide valg, både for seg selv og som medborgere. Derfor er oppgaver i kartleggingsverktøyet knyttet til å vurdere autentiske informasjonskilder, hentet fra nettsider. Elever som møter tekster med motstridende informasjon kan utvikle et mer avansert kunnskapssyn, som vil si at de oppfatter kunnskapene om temaet som mindre sikker og nyansene i saken som viktigere, enn elever som ikke leser om motstridende meninger (Ferguson & Bråten, 2013). En del av ferdighetstesten tar derfor utgangspunkt i to



motstridende tekster, hvor elevene skal vurdere troverdigheten til informasjonen som fremlegges i tekstene. Elevenes ferdigheter til å identifisere og vurdere argumenter blir målt gjennom at elevene må forholde seg til argumentene som fremlegges i de to tekstene, og eksplisitte oppgaver hvor de vil vurdere gyldigheten og relevansen til en rekke argumenter.

### 2.1.3 Hvorfor tenker vi ikke alltid kritisk?

Hvis man skal forstå hvordan elevenes kritiske tenkning fungerer, og hvordan denne kan styrkes, er det relevant å spørre hvorfor vi ikke alltid oppfører oss som kritiske tenkere. En forklaring kan man finne i teorier fra atferdspsykologi, som viser til at vi har to ulike kognitive systemer for å ta avgjørelser (Facione, 2015; Gigerenzer et al., 1999). Et raskt, som tar avgjørelser automatisk, og et tregt, som er kontrollert og bevisst. Det raske systemet er nyttig når man må ta avgjørelser, men ikke har tid til å undersøke og analysere alle aspekter og konsekvenser rundt avgjørelsen. I slike tilfeller er det raske systemet kapabelt til å ta relativt gode avgjørelser basert på lite informasjon (Gigerenzer et al., 1999). Dette systemet anvender *heuristisk tankegang*. *Heuristisk tankegang* kan forklares som kognitive «snarveier» til løsninger på utfordringer (Walton, 2010). Eksempler på dette kan være at vi ofte velger kjente alternativer fremfor det ukjente, uten å undersøke det ukjente alternativet. I tillegg har vi mennesker en tendens til å se etter kausale sammenhenger, selv på et svakt empirisk grunnlag (Walton, 2010). Dette er egenskaper som har blitt utviklet gjennom evolusjonen, og har fortsatt i dag en viktig tidsbesparende virkning på hverdagslige valg (Walton, 2010). Dette systemet anvender alle daglig, og selv filosofen Descartes, som er kjent for sin metodiske tvil, mente at det ikke nyttet å tvile på enhver påstand i hverdagslivet (Gaukroger, 1995)<sup>2</sup>. I mange situasjoner har vi utilstrekkelig med informasjon å belage beslutningene våre på, og ikke tid til å undersøke problemstillingen nærmere. I slike situasjoner er det raske systemet et praktisk

---

<sup>2</sup> Descartes regnes som den første moderne filosof, og hadde som sitt store filosofiske prosjekt å danne et metodisk grunnlag for kunnskap som måtte være sann, foruten enhver tvil (Gaukroger, 1995). Kritisk tenkning skulle være selve garantien for sannhet. Descartes var svært kritisk til sanseerfaringer, og hans metode er et forsøk på å bruke logisk tenkning på å finne påstander som per definisjon er sanne. Mest kjent er han for å vise til at påstanden "Cognito, ergo sum", altså "Jeg tenker, derfor er jeg", er det eneste man vite foruten enhver tvil, fordi det å tvile er per definisjon å tenke, og hvis du tenker er du per definisjon noe som tenker, altså må en som tenker eksistere (Gaukroger, 1995). Descartes mente likevel at hans metode for finne sannheter egnert seg for å finne filosofiske sannheter, men dårlig for å vurdere påstander i hverdagslivet.

og godt verktøy for å ta avgjørelser, basert på erfaring og enkle generaliseringer. I de fleste yrker er dette en viktig egenskap for å være effektiv i jobben sin (Gigerenzer et al., 1999).

Det raske systemet er et evolusjonsmessig gammelt system, og er ikke skikket til å ta avgjørelser som krever planlegging eller omfattende analyser av informasjon, bevis eller konsekvenser (Gigerenzer et al., 1999). Avgjørelser tatt av det raske systemet kan være preget av kognitive bias, fordommer og kan forføres av logisk ugyldige eller irrelevante argumenter. Det raske systemet er derfor lite egnet for kritisk tenkning (Facione, 2015).

Ifølge Walton (2010) er heuristisk tankegang årsaken til at fallasier fremstår som bedre argumenter enn de egentlig er, og at mange lar seg overbevise av dem. Det er altså det raske kognitive systemet som vurderer argumentet, i stedet for det trege. Resultatet blir at avgjørelser som burde vært nøye tenkt gjennom, ender opp som et uberettiget sprang til konklusjon, hvor uten nødvendige analyser av forutsetninger, antakelser og unntak for premissene i argumentet. Etersom fallasier kan overbevise oss til å ta avgjørelser som er dårlig begrunnet, er de farlige (Walton, 2010).

De kognitive ferdighetene, fra Tabell 2-2, som er nødvendige for å tenke kritisk, arbeider i det trege systemet. Det trege systemet kan gjennomføre analyser av ukjente situasjoner og problemstillinger (Facione, 2015). Dette systemet er kapabelt til å planlegge, samle og prosessere informasjon, stille kritiske spørsmål og argumentere for og imot (Gigerenzer et al., 1999). Som vist over kan kritisk tenkning ses på som kompetansen som er nødvendig for å kunne ta veloverveide og reflekterte avgjørelser, og slik tenkning skjer altså i det trege systemet (Facione, 2015), som det vises i Tabell 2-3.

Tabell 2-3: Oversikt over det raske og trege systemet.

	<b>Raske systemet</b>	<b>Trege systemet</b>
<b>Tar avgjørelser ved å:</b>	Hoppe instinktivt til konklusjoner, basert på erfaring og heuristiske snarveier.	Bruke tid til å samle og prosessere informasjon.
<b>Anvendes når:</b>	Det er lite tid eller informasjon tilgjengelig, eller konsekvensene av avgjørelsen er liten.	Man skal ta stilling noe nytt, det er tid og informasjon tilgjengelig og det er viktig å stille kritiske spørsmål.

Både det raske og det trege systemet er påvirket av hverandre. Det raske systemet kan blokkere det trege systemet, hvis personen føler seg stresset eller presset, men det kan også hente kunnskaper som tidligere er prosessert gjennom det trege systemet. På den måten kan det trege systemet bidra til at avgjørelser som blir tatt i det raske systemet er bedre begrunnet.

Facione (2015) mener at den beste måten å styrke elevenes kompetanse til å tenke kritisk er å styrke det trege systemet. Dette begrunnes med at det trege systemet kan læres til å overkjøre det raske systemet, hvis man er tilvendt å tenke kritisk. Hvis elevene møter situasjoner hvor de må anvende det trege systemet, og tenke kritisk, er det mer sannsynlig at de vil gjøre det samme i lignende situasjoner senere.

## 2.2 Relevante oppgavetyper for kartleggingsverktøyet

Kritisk tenkning er et teoretisk konsept, og kan ikke observeres direkte. Innen det kvantitative forskningsfeltet er det vanlig å kalle slike teoretiske konsepter for *begreper*. Begreper som kritisk tenkning må *operasjonaliseres* i kartleggingsverktøyet, før det kan måles. Spørsmålene og oppgavene som måler og tallfester begrepene omtales som *variabler* (Nardi, 2018). Den informasjonen som produseres gjennom variabelen kalles en *verdi*, og er ofte tallfestet. En *sammensatt variabel* er en form for kategori som kan bestå av flere spørsmål (Ringdal, 2018). Utfordringen med å operasjonalisere et teoretisk begrep er å utvikle variabler som gir et treffsikkert mål på begrepet. I en test er variablene formulert som oppgaver som respondenten skal løse. Hva slags type oppgave, og hvordan den er formulert, vil i stor grad påvirke resultatene. Ifølge Haladyna et al. (2002) er det viktig å følge anbefalte retningslinjer når man utvikler oppgaver, hvis man ønsker å oppnå troverdige resultater av høy kvalitet. Dette delkapittelet har som formål å gi innsikt i ulike oppgavetyper som var relevante i utviklingen av ferdighetstesten i kartleggingsverktøyet, for å måle de kognitive ferdighetene som er viktige for å være en kritisk tenker. Det finnes to hovedkategorier av oppgaver som kan brukes i en skriftlig test; *lukkede* og *åpne* oppgaver. Ferdighetstesten er i hovedsak basert på lukkede oppgaver, men i studie 1 anvendes flere åpne oppgaver av typen kortsvarsoppgaver, for å få innsikt i elevenes refleksjon. Ku (2009) anbefaler å anvende noen åpne oppgaver i tillegg til lukkede oppgaver når man måler kritisk tenkning, hvis man ønsker å måle hele begrepet. I det endelige kartleggingsverktøyet var det derfor et mål at det kun skulle brukes kortsvarsoppgaver for å måle de delene av begrepet kritisk tenkning som krever at elevene formulerer egne tanker. Eksempler på dette er delferdighetene «*formulere alternativer*» og «*rettferdiggjøre prosedyrer*», fra Tabell 2-2.

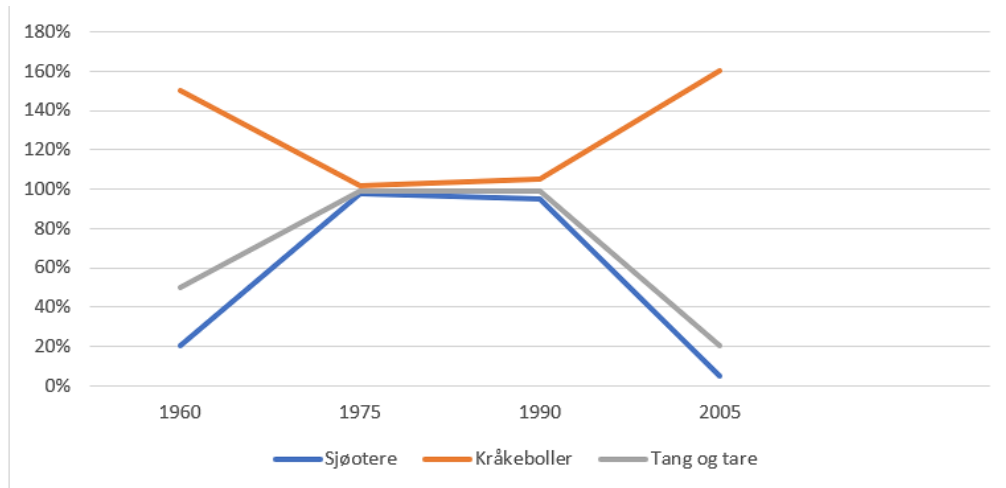
### 2.2.1 Lukkede oppgaver

Lukkede oppgaver er oppgaver hvor respondenten ikke skal formulere et svar, men velger blant ferdigformulerte alternativer (Ringdal, 2018). Lukkede oppgaver er mye brukt i

internasjonale undersøkelser som for eksempel PISA (Sjøberg, 1986). En lukket oppgave består som av en *stamme*, *nøkkel* og en eller flere *distraktorer*. I tillegg kan oppgavene inneholde en *stimulus*.

Stammen består av selve oppgaven som respondenten skal løse, og kan være formulert som et uavsluttet utsagn eller et spørsmål. Haladyna et al. (2002) anbefaler at stammen kun bør inneholde nødvendig informasjon, og være så kortfattet som mulig. For å unngå at respondenten forvirres av oppgaveformuleringen bør den også være tydelig, med lite rom for misforståelser. Man bør derfor unngå begreper som lett kan overses når respondenten leser oppgaven, men som er viktige for å besvare den. Eksempler på dette kan være negasjoner som *aldri* og *ikke* (Patel et al., 1995; Sirnes, 2005).

Det riktige alternativet til oppgaven kalles *nøkkel*, og de alternativene som ikke er riktige kalles for *distraktorer*. Ulike former for lukkede oppgaver har ulike antall nøkler og distraktorer. Ett stimulus kan være en innledende opplysning i forkant av spørsmålet, som for eksempel en illustrasjon, graf eller faktasetning. Figur 2-2 er et eksempel på en slik stimulus som brukes i kartleggingsverktøyet, og medfølgende oppgave.



Figur 2-2: Graf hentet fra kartleggingsverktøyet, hvor den ble anvendt som en stimulus til oppgave 2.5 b).

Eksempeloppgave 1 som er hentet fra kartleggingsverktøyet viser hvordan en oppgave kan være koblet sammen med en stimulus. I oppgaven må elevene først tolke informasjonen om hvordan populasjonen til tre arter utvikler seg over tid i et økosystem fra grafen i Figur 2-2. Elevene må videre lese stammen i Eksempeloppgave 1, som er oppgaveteksten «2.5 b) Hva kan tolkes utfra grafen i figur 1?». Elevene blir presentert for fire svaralternativer, hvor det

andre svaralternativet er nøkkelen, og resten av svaralternativene er distraktorer.

2.5 b) Hva kan tolkes ut fra grafen i figur 1? \*

- Tabellen viser at når det blir flere kråkeboller, blir det flere sjøotere og mindre tang og tare.
- Tabellen viser at nedgangen i antallet sjøotere, skjer samtidig som det blir flere kråkeboller og mindre tang og tare.
- Tabellen viser at når det blir mindre tang og tare, fører dette til at det blir det flere kråkeboller.
- Tabellen viser at når det blir færre kråkeboller, fører dette til at det blir flere sjøotere.

*Eksempeloppgave 1: Eksempel på hvordan en lukket oppgave kan være bygget opp av en stamme, og tre distraktor og nøkkelen til oppgaven (svaralternativ 2).*

Det finnes ulike typer lukkede oppgaver, og de jeg anser som mest relevante for kartleggingsverktøyet er *konvensjonelle flervalgoppgaver (MC)*, *komplekse flervalgoppgaver (CMC)* og *Context dependent Item Set (CDIS)*.

Den konvensjonelle versjonen av flervalgoppgaver (MC), med ett riktig svar, er den mest anvendte formen for lukkede oppgaver, og kan brukes til de fleste kunnskapsområder (Haladyna et al., 2002). Denne formen for lukket oppgave har kun en nøkkel, og flere distraktorer. Downing (2006) argumenterer for at denne typen flervalgoppgaver potensielt kan egne seg til å måle alle nivåer av kompetanse. Oppgaver med kun to svaralternativer kalles *Alternative Choice (AC)*, og kan anses som en avart av flervalgoppgaven. Lukkede oppgaver har ofte blitt kritisert for å være mulige å gjette seg til svaret, og at oppgaver med bare to alternativer er spesielt utsatt for denne feilkilden (Haladyna et al., 2002). Gjetting på AC-oppgaver fører til at vanskelighetsgraden i testen blir lavere enn hvis man har flere distraktorer. For å motvirke denne utfordringen foreslår Downing (2006) å ha mange oppgaver i testen, og å eventuelt stille et krav om at respondenten må score riktig på minst halvparten av oppgavene for å bestå testen.

En annen form for lukket oppgave kalles *komplekse flervalgoppgaver (CMC)*. Disse skiller seg fra konvensjonelle flervalgoppgaver ved at det er flere nøkler. Haladyna et al. (2002) anbefaler å ikke anvende CMC fordi de tar lenger tid å gjennomføre. I tillegg krever oppgaveformen at den som vurderer oppgaven tar stilling til hvordan respondentene tildeles

poeng. Enten må respondentene svare riktig på alle nøklene for å få poeng, eller det kan gis poeng for hver riktig nøkkel, samt minuspoeng hvis respondenten velger en distraktor, slik at man unngår at gjetting blir en betydelig feilkilde. Et mulig argument for å bruke CMC er at de ofte har en høyere vanskelighetsgrad, og det er vanskeligere å gjette seg frem til riktig svar (Kubinger et al., 2010). Denne oppgaveformen ble anvendt i studie 1, men innen studie 2 ble alle de komplekse flervalgsoppgavene delt opp i flere, men mindre oppgaver, for å unngå utfordringene med CMC.

Context dependent Item Set (CDIS) egner seg for å undersøke respondentenes evner til for eksempel problemløsning (Haladyna et al., 2002). En slik oppgave består av innledende informasjon, altså en stimulus, og flere påfølgende oppgaver som er relaterte til stimulusen. Styrken til CDIS-oppgaver er at man kan følge tankerekken til respondenten, slik at lukkede oppgaver kan brukes for å undersøke kompliserte problemstillinger. Denne formen for oppgave er ikke uten ulemper: Den tar ofte mer tid å svare på og det er komplisert å utvikle og vurdere slike oppgaver. I tillegg er det utfordrende at det kan forekomme «følgefeil» hvis respondentene misforstår eller tar feil tidlig i oppgavesettet. Denne typen oppgave brukes i kartleggingsverktøyet, for å kunne anvende autentiske stimuluser som elevene må forholde seg til. Alle de tre formene for lukkede oppgaver som er nevnt over kan være CDIS-oppgaver, hvis de er en del av et oppgavesett koblet til en stimulus.

### 2.2.2 Åpne oppgaver

Det som skiller åpne fra lukkede oppgaver er at åpne oppgaver ikke har ferdigformulerte alternativer. Det er opp til respondenten å bruke egne ord for å formulere svarene (Ringdal, 2018). I hovedsak består åpne oppgaver kun av en stamme, og en eventuell stimulus. Det finnes to hovedtyper av åpne oppgaver, og forskjellen mellom disse er forventningen om lengden på respondentens besvarelse. Den typen av åpen oppgave som er mest relevant for dette studiet er *kortsvarsoppgaver*. En kortsvarsoppgave kan være å be respondenten om å sette navn på, gjøre rede for eller forklare noe, enten knyttet til en stimulus eller ut fra hva respondenten kan fra før. *Langsvarsoppgaver* er den andre formen for åpen oppgave, og kan egne seg for å undersøke sammensatte ferdigheter. Jeg har valgt å utelukke langsvarsoppgaver fra min studie. Årsaken er at den er svært tidskrevende, utfordrende å vurdere og har svært lav reliabilitet sammenlignet med andre oppgaver (McAlpine, 2002).

## 2.3 Klassisk testteori

Det kan være utfordrende å utvikle en test som skal kartlegge et kunnskapsområde eller en ferdighet. Man ønsker å måle den virkelige kompetansen til respondenten, men det kan være mye som påvirker et testresultat, fra om oppgavene undersøker hele kompetansen eller bare deler av det, til det psykologiske presset som kan påvirke respondenten i en «prøvesituasjon» (Cohen et al., 2007). Klassisk test teori er et populært rammeverk for å utvikle tester innen psykologi- og utdanningsfeltet, fordi den tar hensyn til at vi ikke kan observere respondentens *sanne* kompetanse, kun det vi observerer at respondenten svarer på testen (Bichi, 2016).

Utgangspunktet for klassisk test teori er at det finnes en *sann score* (T) for respondenten, som uttrykker respondentens virkelige kompetanse og kunnskap. Utfordringen er at det i virkeligheten vil være feilkilder (E) (feil/errors) som påvirker den observerte scoren (X) (Bichi, 2016). I klassisk test teori er dette formulert i formelen:

$$X = T + E$$

Altså observert score kan uttrykkes som summen av denne sanne scoren og scoren fra feilkildene (Bichi, 2016; Carmines & Zeller, 1979).

Feilkildene som hindrer oss fra å observere respondentens sanne kompetanse kan være både tilfeldige og systematiske (Ringdal, 2018). Systematiske feilkilder gir utfordringer med validiteten i testen (Ringdal, 2018). Det vil si at testen ikke måler den kompetansen man ønsker å undersøke. Feilkildene kan også være tilfeldige, som vil si at respondentens score ikke er reliabel. Testscoren blir da et unøyaktig mål på den kompetansen man ønsker å undersøke (Ringdal, 2018). I dette delkapittelet presenteres de aspektene som er viktige for å kvalitetssikre en test, for å få innsikt i respondentens sanne kompetanse, og minske feilkildene. Det vil redegjøres for hvordan reliabilitet og validitet kan undersøkes i en test, og hvordan testens og oppgavens vanskelighetsgrad og evne til å diskriminere kan bidra til å få innsikt i respondentenes evner. Slike vurderinger av oppgaver eller andre variabler kalles *item-analyse*. I dette masterprosjektet har jeg valgt å anvende begrepet *kvalitetsmarkører* som et samlebegrep for alle mål som kan gi innsikt i feilkildenes påvirkning på testen.

### 2.3.1 Kriteriebasert test

Det finnes ulike typer tester, og hva slags type test man utvikler er påvirket av kompetansen man ønsker å måle. I tidligere delkapitler har jeg gjort rede for kjennetegn på kompetanse som

er nødvendig for å tenke kritisk i naturfag. Kjennetegnene kan ses på som en form for kriterier, som blir utgangspunktet for kartleggingsverktøyet, og er derfor en *kriteriebasert test*. En kriteriebasert test undersøker hva respondentene mestrer innen et ferdighets- eller kunnskapsområde, ut fra kriterier som kjennetegner god mestring (Cohen et al., 2007). En kriteriebasert test trenger ikke å være *parametrisk*, i motsetning til en normbasert test. Normbaserte tester har som hensikt å være parametriske, altså ha en normalfordeling av verdier. Normbaserte tester tar derfor utgangspunkt i gjennomsnittet i populasjonen, og en respondents score forteller oss hvordan respondenten gjorde det i forhold til gjennomsnittet (Cohen et al., 2007). IQ-tester er et eksempel på en slik test, hvor gjennomsnittsscoren for hele populasjonen skal være 100 (Farmer et al., 2020). Slike tester er svært utfordrende å utvikle, og svært avhengig av å prøves ut i et representativt utvalg (Cohen et al., 2007). Selv om ferdighetstesten er en kriteriebasert test, vil jeg undersøke om resultatene på testen følger en normalfordeling.

### 2.3.2 Testens vanskelighetsgrad

Både oppgavers vanskelighetsgrad og diskrimineringssevne kan først tallfestes når en test er utprøvd på respondentene. Man kan regne ut *item facility* ( $p$ ) for å tallfeste hvor vanskelig en oppgave er. Hvis en oppgave har høy *p-verdi* vil de fleste respondentene få til oppgaven, mens oppgaver med lave verdier vil det være få som klarer å svare riktig på (McAlpine, 2002). En oppgave med høy vanskelighetsgrad vil altså ha lave  $p$ -verdier, og en oppgave med høye  $p$ -verdier vil ha en lav vanskelighetsgrad. Item facility er et mål på andelen som fikk til oppgaven, og kan omgjøres til prosent ved å multiplisere  $p$ -verdien med 100 (Cohen et al., 2007). Derfor vil denne oppgaven utrykke vanskelighetsgrad gjennom  $p$ -verdi og som prosent der det er hensiktsmessig. Item facility er uttrykt gjennom følgende formel:

$$\text{Item facility } (p) = X_g/X_{\max}$$

$$X_g = \text{Gjennomsnittscore på oppgaven}$$

$$X_{\max} = \text{Den maksimale poengsummen til oppgaven.}$$

Hvis målet med en test er å få et så godt overblikk over respondentenes kompetanse, bør man ifølge McAlpine (2002) ta hensyn til testens vanskelighetsgrad. Hvis hele testen er for vanskelig for utvalget vil man ikke få innsikt i den kompetansen respondentene innehar, fordi de i liten grad har fått vist den. Respondentene i et utvalg vil ha ulike grad av kompetanse i



kunnskapsområdet som testes, men hvis alle oppgavene er for vanskelig for hele utvalget, vil man ikke klare å skille mellom deres kompetansenivå. På samme måte vil en test som er for lett, ikke klare å skille mellom respondentenes kompetanse, fordi alle fikk til oppgavene.

En tommelfingerregel er at den optimale gjennomsnittlig p-verdien for alle oppgavene i en test bør være ca. 0.5 for utvalget, hvis man ønsker å oppnå maksimal differensiering (McAlpine, 2002). I tester hvor det er mange oppgaver som gir liten uttelling av poeng er det viktig å variere vanskelighetsgraden. Hvis det er god variasjon av vanskelighetsgrader i oppgavene vil respondentene klare de oppgavene som er lette nok for dem, men ikke de som er for vanskelige. Respondenter med lav kompetanse, i forhold til resten av utvalget, vil da kun klare de oppgavene med høy p-verdi, mens respondentene med høy kompetanse vil klare nesten alle oppgavene. Hvis for eksempel p-verdien til en oppgave er på 0.30, som altså er en oppgave med høy vanskelighetsgrad, bør man ha en tilsvarende lett oppgave, med 0.70 i p-verdi, slik at snittverdien blir 0.5. Dette er spesielt viktig for normbaserte tester, siden slike tar utgangspunkt i å skille respondentene fra hverandre i størst mulig grad (Cohen et al., 2007). Kriteriebaserte tester må ikke nødvendigvis ha en fastsatt gjennomsnittlig vanskelighetsgrad som mål, fordi kompetansen eller kriteriet som undersøkes kan være i seg selv enkel eller utfordrende (Cohen et al., 2007). Kriteriebaserte tester bør søke å ha en vanskelighetsgrad i oppgavene som tilsvarer hvor vanskelig kompetansen som man måler er (Cohen et al., 2007). Testen i kartleggingsverktøyet er en kriteriebasert test, og den gjennomsnittlige p-verdien kan være enten høyere eller lavere ut fra om flertallet av respondentene innehar ferdighetene som undersøkes eller ikke. Likevel er det oppgaver som skiller godt mellom respondentenes kompetanse som er interessante, og derfor vil vanskelighetsgraden til oppgavene undersøkes og anvendes som et kriterium for å forkaste ikke-fungerende oppgaver.

Det er ikke bare testens gjennomsnittlige vanskelighetsgrad som er viktig, for man bør undersøke vanskelighetsgraden til hver enkelt oppgave. Oppgaver med høyere p-verdi enn 0.85 eller lavere enn 0.15 bør som regel unngås, fordi oppgavene bidrar lite med å skille kompetansenivået til respondentene (McAlpine, 2002). Årsaken er at nesten ingen vil klare oppgaver hvor  $p=0.15$ , og så godt som alle vil klare oppgaver med en p-verdi på over 0.85. Unntak fra dette kan tenkes om man har et stort utvalg og mulighet til å gjennomføre en omfattende test (McAlpine, 2002). Cohen et al. (2007) anbefaler å bruke oppgaver som har en vanskelighetsgrad mellom 0.33 og 0.67, men dette bør man vurdere opp mot andre hensyn, som hvor mye oppgaven bidrar til testens innholdsvaliditet. I denne masterstudien på et

begrenset utvalg vil det være et mål at flertallet av oppgavene holde seg innenfor rammen på 0.15 og 0.85, og helst mellom 0.33 og 0.67.

### 2.3.3 Testens diskrimineringssevne

Hensikten med en oppgave er å kunne måle en kompetanse, og at kun de som innehar denne kompetansen skal klare oppgaven. Dette er ikke alltid tilfellet, og i så fall *diskriminerer* ikke oppgaven godt nok. Diskrimineringssevne ( $d$ ) er et mål på om de som får til en oppgaven er de som innehar kompetansen man ønsker å måle (Cohen et al., 2007). Dette er en viktig faktor å ta hensyn til under testutvikling, og det er ønskelig at oppgavene har god evne til å diskriminere (McAlpine, 2002). Oppgavens diskrimineringssevne gis ved å undersøke korrelasjonen mellom gjennomsnittscoren på hele testen og scoren på oppgaven som undersøkes. For å undersøke korrelasjonen kan man anvende Pearsons produktmomentkorrelasjon-koeffisienten. Den kan brukes på tester hvor hele testen undersøker samme kompetanse (McAlpine, 2002). Eventuelt kan den anvendes i tilfeller hvor bare deler av testen undersøker den samme kompetansen. Da bør man lage en egen testkategori med oppgaver som måler den samme kompetansen, og undersøke diskrimineringssevnen til oppgavene i forhold til testkategorien. På denne måten kan oppgavens diskrimineringssevne tallfestes som en verdi mellom +1, ved full korrelasjon mellom oppgaven og resten av testen, og -1, hvis det er en omvendt korrelasjonssammenheng (Cohen et al., 2007). Man bør ikke oppnå en negativ verdi, hvis oppgavene måler samme kompetanse. Oppgaven diskriminerer bedre desto nærmere verdien er +1 (McAlpine, 2002).

McAlpine (2002) anbefaler å anvende oppgaver med diskrimineringssevne høyere enn 0,3, og beskriver en diskrimineringssevne på 0,4 som svært god. Lie et al. (2005) anvender også en nedre grense for diskrimineringssevne på 0,3, og forklarer denne grensen ved at oppgaver med høyere diskrimineringssevne enn 0,3 bidrar positivt til å øke testens reliabilitet. Ifølge McAlpine (2002) bør oppgaver som diskriminerer under 0,3 undersøkes nærmere. Man kan undersøke om oppgaveformuleringen kan være årsaken til at enkelte oppgaver har liten evne til å diskriminere (ledende svar, kan oppgaven tolkes på andre måter osv.). Problemet kan også ha sammenheng med at det er vanskelig å få høye verdier for diskriminering hvis oppgavene har for lav vanskelighetsgrad (Lie et al., 2005). Å endre på eller å fjerne oppgaver er alltid et vurderingsspørsmål for testutvikleren (Cohen et al., 2007). I denne masteroppgaven vil 0,3 anses som akseptabel diskrimineringssevne, men det er ønskelig med høyere verdier.

Hvis en oppgave har lav diskrimineringssevne, kan feilen være at distraktorene ikke fungerer. Effektiviteten til hver distraktor kan undersøkes ved å se hvor mange ikke-kompetente og kompetente respondenter som velger distraktoren (Cohen et al., 2007). En distraktor er effektiv hvis den «lokker» elever som mangler ferdighetene som måles til å velge den, men ikke de kompetente elevene. Elevene som anses som kompetente er altså de som har fått til mange oppgaver i testen som helhet eller i testkategorien oppgaven tilhører. En enkel måte å undersøke effektiviteten av en distraktor er ved å undersøke hvor godt oppgaven diskriminerer, og hvis den diskriminerer godt, hvilke distraktorer ble valgt. Hvis oppgaven har en lav diskrimineringssevne, kan en mulig forklaring være at de kompetente elevene velger en av distraktorene. Da bør man vurdere å omformulere eller fjerne distraktoren. Ifølge Sirnes (2005) er det utfordrende og tidskrevende å utarbeide sannsynlige distraktorer, og man bør ikke undervurdere arbeidet som kreves for å utvikle gode testverktøy basert på lukkede oppgaver.

#### 2.3.4 Reliabilitet i testutvikling

En tests reliabilitet handler om hvor konsekvente målingene som innhentes er, altså hvor stor del av målingene som skyldes tilfeldige feil. Ifølge Patel et al. (1995) kan reliabilitet oppsummeres som påliteligheten til målingene. I en test med høy reliabilitet vil det være lite innvirkning av tilfeldige feilkilder som påvirker den scoren, altså den observerte kompetansen til respondenten (Patel et al., 1995). En form for allmenn kildekritikk er den enkleste måten å vurdere reliabilitet til et måleinstrument på. Gjennom å undersøke hvordan spørsmålene er formulert kan vi spore mulige feilkilder, og denne metoden vil være viktig i utformingen av kartleggingsverktøyet. Under utvikling av spørreskjemaer eller tester er denne formen for vurdering av reliabilitet svært viktig (Ringdal, 2018).

Man kan tallfeste reliabiliteten til en test ved å sammenligne den med en velutprøvd test som måler det samme (*Parallell forms*), eller gjennomføre en test to ganger på samme utvalg (*test-retest*) (McAlpine, 2002). Utfordringer med begge metodene er at de er krevende å få til når det er begrenset med tid og ressurser tilgjengelig. McAlpine (2002) mener derfor at det i mange tilfeller er fordelaktig å anvende en tredje metode for å måle reliabilitet: *Intern konsistens*. Metoden går ut på å undersøke korrelasjonen mellom oppgavene i en test som er ment for å måle den samme kompetansen (Carmines & Zeller, 1979; Sirnes, 2005). Å måle intern konsistens krever kun én gjennomføring av testen (McAlpine, 2002), og passer når man anvender et *tverrsnittdesign*, altså måler kompetansen til respondentene ved kun ett tilfelle

(Ringdal, 2018). Den vanligste metoden for å tallfeste intern konsistens er *Cronbach alfa* ( $\alpha$ ), og det er viktig mål innen testutvikling (Cohen et al., 2007). Derfor vil jeg anvende Cronbach alfa som en kvalitetsmarkør i item-analysen gjennom testutviklingen, sammen med item-facility og diskrimineringsevne.

Med Cronbach alfa undersøker man om det er korrelasjon mellom flere variabler som skal måle det samme, og hvis det er en sterk korrelasjon har variablene en intern konsistens. En høy verdi av Cronbach alfa indikerer at reliabiliteten er god (Cohen et al., 2007). Cronbach alfa er en statistisk størrelse som varierer fra 0 til 1. Jo nærmere 1 verdien er, jo mer intern konsistens er det (Friborg, 2010). Størrelsen på Cronbach alfa påvirkes av hvor mange variabler som måler det samme, og hvor sterk korrelasjonen mellom disse er (Eikemo & Clausen, 2012). Taber (2018) understreker at veldig høye  $\alpha$ -verdier ikke alltid er ønskelig, hvis ikke måleinstrumentet måler et smalt og endimensjonalt forskningsobjekt. Tavakol og Dennick (2011) anbefaler derfor at  $\alpha$ -verdier ikke overstiger 0.9 hvis man måler et heterogent og flerdimensjonalt begrep. Så høye  $\alpha$ -verdier indikerer som regel at man har overflødige oppgaver, og omfanget av testen bør reduseres.

Selv om alfa er det mest anvendte målet for reliabilitet i testutvikling advarer Cronbach selv om at man ikke bør tillegge målet for mye verdi, og argumenterer for at *standardfeil* gir et bedre helhetsinntrykk av hvor reliabelt et måleinstrument er (Cronbach & Shavelson, 2004). For eksempel kan måleinstrumenter med få variabler eller tester med mange AC-oppgaver gi lave  $\alpha$ -verdier, og dette betyr ikke nødvendigvis at målene er upresise (Wibowo, 2016). *Standardfeil* kan utregnes hvis man kjenner til standardavviket og har et reliabilitetsmål for testen. Standardfeilen er et mål på hvor stor feilmargin man kan forvente i testen (McAlpine, 2002). Det vil si at man kan tallfeste den gjennomsnittlige avstanden fra sann score til scoren i testen, dersom testen er en valid operasjonalisering av begrepet som skal måles.

En annen relevant måte å undersøke reliabilitet handler om i hvilken grad den som vurderer testen påvirker resultatene. Dette omtales som *sensorreliabilitet*, og kan utregnes ved å undersøke om hvor stor korrelasjonen er mellom de som vurderer testen (Nardi, 2018). For å øke denne formen for reliabilitet kan man øke antallet oppgaver og å ha klare vurderingskriterier. For å oppnå optimal sensorreliabilitet kan man anvende selvrettende lukkede oppgaver med fasit.

### 2.3.5 Validitet i testutvikling

Validitet handler om man faktisk undersøker det vi tok sikte på å undersøke (Patel et al., 1995). Validitet i forbindelse med testutvikling kan oppsummeres om testen måler det den er ment å måle (Cohen et al., 2007). Hensikten med testen er å kunne måle ferdigheter som er viktige for kritisk tenkning, og måler ikke oppgavene denne kompetansen har testen lav validitet. *“If a piece of study is invalid then it is worthless.”*(Cohen et al., 2007, s. 133).

Validitet må ses i sammenheng med reliabilitet (Patel et al., 1995). Hvis et måleinstrument har lav reliabilitet, vil det også ha lav validitet (Ringdal, 2018). Årsaken er at målinger som ikke er pålitelige heller ikke måler det vi ønsker å måle. Høy reliabilitet er derimot ingen garanti for høy validitet.

Det finnes ulike former for validitet som er relevante for testutvikling. I en masteroppgave hvor målet er å operasjonalisere begrepet kritisk tenkning blir *begrepsvaliditet* svært viktig. En høy begrepsvaliditet er et resultat på hvor godt samsvar det er mellom oppgavene og det teoretiske begrepet (Kleven & Hjordemaal, 2018). For å oppnå dette er det viktig at alle delene av begrepet blir målt like mye, og at ikke oppgavene måler forstyrrende og irrelevante elementer som ikke har noe med begrepet man ønsker å undersøke å gjøre (Kleven & Hjordemaal, 2018). Det finnes ulike måter å undersøke begrepsvaliditeten på. Ifølge Nardi (2018) kan testutviklerne komme et stykke på vei ved å bruke sunn fornuft, og stille spørsmål om det virker fornuftig at oppgaven faktisk måler det du ønsker å måle. Dette kalles *umiddelbar validitet*. *Innholdsvaliditet* er en annen form for validitet som handler om hvor representative oppgavene er for det kunnskapsinnholdet som testen er ment for å måle (Sirnes, 2005). Det er sjeldent at et teoretisk begrep operasjonaliseres til målbare verdier uten å miste noe av den teoretiske rikdommen sin. Innholdsvaliditet handler også om hvordan de ulike elementene er vektet i poengutdelingen ved testscore. Man kan undersøke innholdsvaliditet gjennom å gjøre en skjønsmessig analyse av innholdet i måleinstrumentet (Patel et al., 1995), altså vurdere om hele begrepet måle dekkes gjennom variablene (Ringdal, 2018). Det er en slik validitetsvurdering som anvendes i dette masterprosjektet. Det finnes kvantitative analyser for å tallfeste validiteten til et måleinstrument, som faktoranalyse, eller ved sammenligne resultatene på testen man vil undersøke med velutprøvd test med god validitet i samme utvalg. Hvis man ønsker å oppnå et mer presist bilde av et måleinstruments validitet, bør skjønsmessige validitetsvurderinger suppleres med slike analyser. På grunn av masterprosjektets omfang, og at validiteten er avhengig av reliabilitet, har jeg valgt å fokusere

mer på å undersøke testens reliabilitet enn validitet. Selv om man tar høyde og hensyn til alt som er nevnt over i testutviklingen, vil det alltid være noe tilfeldige og systematiske feil i alle målinger (Kleven & Hjordemaal, 2018). Derfor kan vi ikke forvente å oppnå perfekt reliabilitet eller validitet, men søke å minimere feilkildene.

### 2.3.6 Valg av oppgavetype med utgangspunkt i reliabilitet og validitet

Hensikten med en test er å få innsikt i den reelle kompetansen en respondent har innenfor et bestemt kunnskapsområde. Dette er som nevnt ikke mulig ifølge klassisk testteori, fordi man kun observerer kompetansen indirekte gjennom oppgavene i testen. Man bør derfor ta hensyn til oppgaveformens validitet og reliabilitet når man bestemmer hvilken oppgavetype som skal anvendes i testen.

Bruk av lukkede oppgaver i et standardisert kartleggingsverktøy har mange fordeler over åpne oppgaver, spesielt med tanke på reliabilitet. McAlpine (2002) argumenterer for at alle tiltak som gjøres for å kvalitetssikre en høy reliabilitet vil potensielt bedre testens validitet. Det stilles ulike forventninger til reliabiliteten til ulike typer oppgaver. Hovedregelen er at desto mer konkret en oppgave er, jo høyere intern konsistens kan man forvente (McAlpine, 2002). Minstekravet for Cronbach alfa i en test blant lukkede oppgaver bør være 0,65. (McAlpine, 2002). Også med hensyn på sensorreliabilitet er det hensiktsmessig å anvende lukkede oppgaver, ettersom dette bør føre til en fullstendig positiv korrelasjon mellom de som evaluerer testen, spesielt ved bruk av selvrettende programvare (Kleven & Hjordemaal, 2018). Åpne oppgaver vil derimot kunne bli svært påvirket av den som evaluerer besvarelsene. Downing (2006) mener at det er så stor forskjell i reliabiliteten til åpne og lukkede oppgaver at man bør anvende lukkede oppgaver hvis det er et alternativ. Samtidig vil det i forbindelse med denne studien være mulig å teste et større utvalg, fordi det vil være mindre tidskrevende å vurdere hver enkelt respondents svar på testen.

En annen fordel med lukkede oppgaver er at respondenten bruker langt mindre tid per oppgave (Kleven & Hjordemaal, 2018). Det muliggjør at testen kan inneholde et høyere antall oppgaver. Verdien på Cronbach Alpha er avhengig av antallet oppgaver, og vil øke med flere oppgaver. Lukhele et al. (1994) gjennomførte en studie hvor lukkede og åpne oppgaver på en test i kjemi ble sammenlignet. I oppgavesettet som ble brukt i studien rakk respondentene gjennomsnittlig å svare på 16 MC-oppgaver på tiden det tok å løse en åpen oppgave. Studien konkluderte med at de åpne oppgavene ikke ga mer informasjon om elevenes

kjemikompetanse enn MC-oppgavene, men testen med de åpne oppgavene tok over dobbelt så lang tid å gjennomføre (Lukhele et al., 1994). Et økt antall oppgaver vil derfor kunne øke validiteten til kartleggingsverktøyet, ved at flere ferdigheter tilknyttet kritisk tenkning kan undersøkes med mange oppgaver med ulik tilnærming i løpet av testsituasjonen. Også reliabiliteten øker som følge av at lukkede oppgaver er mindre tidskrevende. Man kan derfor ha flere oppgaver i testen, noe som styrker den interne konsistensen.

En vanlig kritikk mot lukkede oppgaver er at det er mulig å gjette på hvilket svar som er riktig, og at gjennom å eliminere usannsynlige distraktorer kan man komme frem til riktig svar (Downing, 2006). Denne utfordringen er ikke til stede ved åpne oppgaver, fordi respondenten må formulere svaralternativene selv. Konsekvensen av gjetting er at det vil virke som respondenten har et større kunnskapsgrunnlag enn realiteten, og vil dermed gi tilfeldige resultater som påvirker testens reliabilitet. Det er noen tiltak som kan motvirke gjetting. For det første er det viktig å kvalitetssjekke testen før den tas i bruk, og fjerne oppgaver med svak diskrimineringssevne. Det enkleste er likevel å øke antallet oppgaver, for selv om det er mulig å gjette på enkeltoppgaver vil det være vanskelig å få til en høy poengsum på testen. Sjansen for å gjette riktig minker med antallet oppgaver i testen og antallet sannsynlige distraktorer i oppgaven. At det er mulig å eliminere distraktorer i en oppgave, og på den måten komme frem til riktig svar, kan motvirkes gjennom å øke antallet *sannsynlige* distraktorer i oppgavene.

Til tross for lukkede oppgavers styrke med hensyn til reliabilitet, er det likevel i de fleste tilfeller viktigere å velge den formen av oppgaver som måler kompetansen man ønsker å undersøke på mest mulig valid måte (McAlpine, 2002). Åpne oppgaver bør anvendes når man for eksempel undersøker elevenes evne til å formulere seg. Samtidig kan lukkede oppgaver anses som mer valide nettopp fordi eleven ikke er avhengig av å være flinke til å formulere seg skriftlig for å kunne svare. Et tradisjonelt syn på testing tilsier at lukkede oppgaver ikke kan måle kompliserte kompetanser på en adekvat måte, og det derfor er hensiktsmessig å anvende åpne oppgaver for dette (Lukhele et al., 1994). I tillegg argumenterer Messick (1993) for at åpne oppgaver kan gi en bredere innsikt hva respondenten kan. Haladyna et al. (2002) argumenterer for at hva som kan måles med lukkede oppgaver kun avhenger av hvor kompetente de som utvikler testen er. For å måle ferdigheter som for eksempel krever at elevene gjøre rede for og formulerer egne tanker mener jeg likevel at lukkede oppgaver er mindre egnet enn åpne oppgaver. Åpne oppgaver vil derfor brukes for å måle de ferdighetene

som krever at studenten formulerer, forklarer eller gjør rede for egne tanker og resonnementer.

## 2.4 Studier på eksisterende tester

Kritisk tenkning har fått et økt fokus i opplæringen gjennom de siste tiårene, og denne oppmerksomheten har ført til et ønske om å kunne teste elevens evne til kritisk tenkning, spesielt i USA (O'Hare, 2005). Til tross for oppmerksomheten finnes det likevel kun en håndfull anerkjente og velutprøvde måleinstrumenter. Man kan skille mellom to hovedtyper tester av kritisk tenkning; generelle og fagspesifikke tester. De fleste testene anvender kun lukkede oppgaver, men det finnes også tester som fokuserer på kortsvarsoppgaver og langsvarsoppgaver (O'Hare, 2005). De mest kjente måleinstrumentene er «Watson-Glaser Critical thinking appraisal» (Watson og Glaser, 1990), «Cornell Critical thinking test» (Ennis og Millman, 1985) og «California Critical thinking skills test» (Facione et al., 1991). Alle de nevnte testene tar utgangspunkt i at kritisk tenkning kan deles inn i ferdigheter og holdninger, og måler kun ferdigheter. Oppgavene i testene undersøker elevens kognitive ferdigheter til å trekke logiske slutninger. Facione har i tillegg utviklet en egen test for å måle studenters holdninger til kritisk tenkning; «The California Critical thinking disposition inventory» (Facione et al., 2000). I manualene til testene oppgis det lite *psykometrisk* informasjon om testene, som det er vanlig å stille til måleinstrumenter innen psykologifeltet, mener O'Hare (2005), som har sammenlignet testene. De kvalitetsmarkørene jeg vil anvende for item-analyse er eksempler på psykometriske data, som forteller noe om kvaliteten til testen.

For eksempel oppgir verken testen fra Watson og Glaser eller Cornell Critical thinking test (level z) mål på reliabiliteten til de enkelte ferdighetene som måles. De oppgir heller ikke mål på validitet, som for eksempel resultater fra faktoranalyse. Testen som Facione har vært med å utvikle, California Critical thinking skills test, er en svært populær test. O'Hare (2005) mener at testens popularitet skyldes at den er direkte basert på ferdighetene slik de ble utformet i delphistudiet til Facione (1990a). I tillegg er det få andre kjente tester i forhold til interessen for kritisk tenkning (O'Hare, 2005). Testen måler kun fem av ferdighetene som ble identifisert i delphistudiet, og det oppgis ikke verdier for intern konsistens for de enkelte ferdighetene i testmanualen (O'Hare, 2005). I et studie gjennomført av Leppa (1997) ble Cronbach alfa for de enkelte ferdighetene målt til å være mellom 0,21 og 0,51. Facione (1990b) rapporter selv at California Critical thinking test har en Cronbach alfa mellom 0,68 til 0,69 for de 34 oppgavene i testen. Reliabilitetsmålene er ikke så høye som man kan forvente av en anerkjent



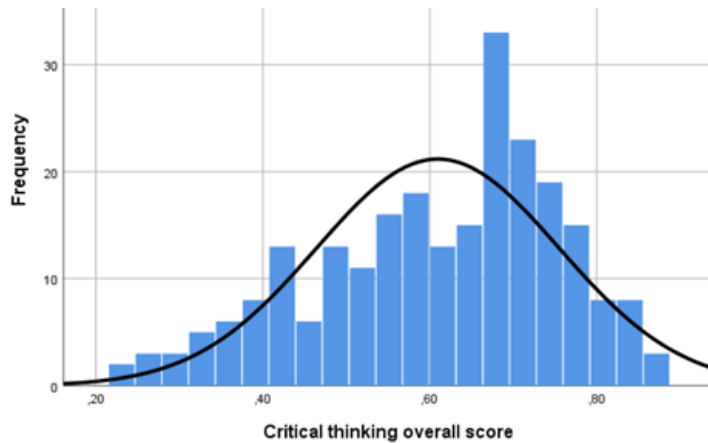
test, men Facione mener selv at årsaken er at mange av oppgavene er dikotome, altså er det anvendt AC-oppgaver, og ettersom testen måler ulike kognitive ferdigheter kan man ikke forvente høy intern konsistens (O'Hare, 2005). Loo og Thorpe (1999) har undersøkt den interne konsistensen for de fem ferdighetene som blir målt i Watson-Glaser Thinking Appraisal test (Form S). Hver ferdighet blir målt med mellom 7-9 oppgaver, og Cronbach alfa for de enkelte ferdighetene varierer mellom 0,17 og 0,74. Flere studier viser til validitetsutfordringer i målene av de enkelte ferdighetene i Watson-Glaser test og California-testen når testresultatene blir undersøkt gjennom faktoranalyse, og dette kan tyde på at ferdighetene er sammenkoblet og avhengige av hverandre (Kluge et al., 2018). Hvis det er tilfellet er det bedre å undersøke score på hele testen enn de enkelte ferdighetene (Kluge et al., 2018). Ifølge O'Hare (2005) virker det som de tre nevnte testene har god begrepsvaliditet sett opp mot den teoretiske diskusjonen om kritisk tenkning. Han konkluderer med at den filosofiske diskusjonen om kritisk tenkning har kommet langt, men at testutviklingen innen temaet fortsatt har ett stykke å gå sett fra et psykometrisk standpunkt.

Det har i de siste tiårene også blitt utviklet fagspesifikke tester for kritisk tenkning i naturfag, slik som jeg har til hensikt å gjøre i dette masterprosjektet. Mapeala og Siew (2015) har gjennomført en pilotstudie på 30 femteklassinger i Malaysia med 55 oppgaver om kritisk tenkning og logisk problemløsning innen naturfaglige temaer. Kun 30 av oppgavene hadde ønskelig vanskelighetsgrad og evne til å diskriminere. Tiruneh et al. (2017) har utviklet en kriteriebasert test som måler kritisk tenkning knyttet opp mot temaet elektrisitet og magnetisme i fysikk på universitetsnivå. Cronbach alfa for testen ble målt til 0,72, og kun 7 av 20 oppgaver diskriminerte på et akseptabelt nivå ( $d > 0,3$ ), med hensyn til at oppgavene skal bidra positivt til den interne konsistensen. Den interne konsistensen er lavere enn forventet, ifølge Tiruneh et al. (2017), og dette kan skyldes at flere ulike kognitive ferdigheter måles. Den gjennomsnittlige vanskelighetsgraden var under 0,5, men dette kan anses som akseptabelt ettersom det var en kriteriebasert test.

Det anbefales å anvende velutprøvde variabler og tester i stedet for å utvikle egne (Ringdal, 2018), ettersom det er utfordrende å utvikle et måleinstrument av høy kvalitet. Jeg har ikke funnet en eneste hel test som er offentlig tilgjengelig. De eksempeloppgavene jeg har undersøkt har gitt inntrykk av at tester innen kritisk tenkning i stor grad fokuserer på logisk resonnering, og at oppgavene kan virke virkelighetsfjerne fra de problemstillingene elevene vil møte i sin hverdag. En test som fokuserer på kritisk tenkning, men med utgangspunkt i

autentiske problemstillinger innen naturfaglige temaer er utviklet gjennom PROBE-prosjektet av Kluge et al. (2018). Testen har blitt utviklet og testet gjennom tre studier, og den endelige versjonen består av 53 spørsmål som har blitt testet på 271 elever på 10. trinn. Det ble anvendt både lukkede og kortsvarsoppgaver, i tråd med Ku (2009) sin anbefaling om å variere oppgavetyper når man måler kritisk tenkning, for å ivareta innholdsvaliditeten knyttet til begrepet. Årsaken er at åpne oppgaver er bedre egnet til å undersøke hvordan elevene begrunner sine valg og strategier for å komme frem til en konklusjon. Dette er en viktig del av det å tenke kritisk.

Testen målte fire ferdigheter og to holdningskategorier. Cronbach alfa for de 39 oppgavene i testen er målt til 0,84, som Kluge et al. (2018) argumenterer for er den høyeste verdien som har blitt rapportert om innen testing av kritisk tenkning. Cronbach alfa for de enkelte ferdighetene er målt er mellom 0,30 til 0,74. Testen er delt i tre deler med fokus på hvert sitt tema, og den interne konsistensen innen hver del er mellom 0,68 og 0,77. Det betyr at det er høyere intern konsistens i testens tre tematiske deler, enn de ferdighetene som ble undersøkt. Det ble gjennomført en faktoranalyse på testen, for å tallfeste validitet med utgangspunkt i om det er mulig å skille ferdighetene ut fra datamaterialet. Faktoranalysen viser at man ikke kan skille mellom oppgaver knyttet til de ulike ferdighetene. Kluge et al. (2018) konkluderer med at ferdighetene sannsynligvis ikke er endimensjonale, og det var større sammenheng mellom oppgaver med samme faglige tema enn de ulike kognitive ferdighetene som ble målt. De oppgavene som undersøkte elevenes evne til å forstå den naturfaglige informasjonen hadde lavest vanskelighetsgrad, mens de oppgavene som undersøkte elevenes evne til å begrunne valgene sine var testens vanskeligste. Kluge et al. (2018) fant ingen sammenheng mellom gjennomføringstid og score i testen. Figur 2-3 gir oversikt over fordelingen av testscore fra PROBE-prosjektet (Kluge et al., 2018). Kluge et al. (2018) argumenterer for å ha oppnådd en tilnærmet normalfordeling av resultatene, med en forskyvning av kurven mot høyre. Det oppgis ingen resultater på kvantitative tester som kunne vært anvendt for å teste normalfordeling.



Figur 2-3: Fordeling av testscore i PROBE-prosjektet (hentet fra Kluge et al. (2018))

Ettersom testen i PROBE-prosjektet har et tilsvarende fokus som jeg har valgt, forsøkte jeg å ta kontakt med testutviklerne via epost, og fikk tilgang til rapporten om prosjektet, men ikke testen. I rapporten er det linket til testen, men nettadressen var ikke i bruk lenger. En svakhet ved å anvende en eksisterende test er at den ikke er spesialtilpasset det kunnskapsområdet man ønsker å undersøke (Cohen et al., 2007). Formålet med masterprosjektet er derfor å utvikle og teste et kartleggingsverktøy for kritisk tenkning i naturfag. De fleste oppgavene i ferdighetstesten i kartleggingsverktøyet er basert på virkelighetsnære problemstillinger, med fokus på kildekritikk av autentiske tekster og å kunne fatte velbegrunnede konklusjoner. PROBE-testen, utviklet av Kluge et al. (2018), har et tilsvarende fokus, og er derfor et godt utgangspunkt for å sammenligne resultatene fra dette masterprosjektet med.

### 3 Metode

For å utvikle, kvalitetssikre og teste ut kartleggingsverktøyet har jeg valgt et forskningsdesign basert på kvantitative metoder. Et forskningsdesign innebærer hva det fokuseres på i studien, hvilken populasjon og utvalg som studeres og hvordan undersøkelsen gjennomføres (Thagaard, 2013). I dette kapittelet redegjøres det for de metodiske valgene i forskningsdesignet. Det gis først en oversikt over hvordan begrepet kritisk tenkning i naturfag ble operasjonalisert i første utkast av kartleggingsverktøyet. Videre gis det en oversikt over prosjektets oppbygning, formålet med de ulike studiene og utvalgene som studeres. Det redegjøres også for hvordan den innsamlede dataen ble bearbeidet og analysert gjennom studiene. I tillegg beskrives etiske avveininger og hensyn som ble tatt i prosjektet. Studienes reliabilitet og validitet diskuteres gjennom masteroppgaven, med utgangspunkt i teori knyttet til testutvikling, som ble gjort rede for i kapittel 2.

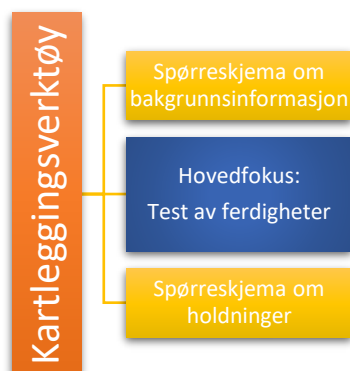
I starten av et forskningsprosjekt er det viktig å avgrense studiets fokus. Fokuset i denne oppgaven er utviklingen og utprøvingen av kartleggingsverktøyet. Innen kvantitativ forskning anbefales det ikke for nybegynnere å falle for fristelsen til å utvikle egne måleinstrumenter (Ringdal, 2018), og testutvikling er en spesielt arbeidskrevende og omfattende forskningsmetode (Cohen et al., 2007). Til tross for dette har jeg valgt å utvikle et eget måleinstrument av to grunner. For det første eksisterer det få tester om kritisk tenkning i naturfag, med fokus på autentiske naturfaglige problemstillinger og kildekritikk. Hvis jeg skulle tatt utgangspunkt i en eksisterende test, er det en svakhet at eksisterende tester ikke er spesialtilpasset det kunnskapsområdet man ønsker å undersøke (Cohen et al., 2007), og den eneste testen som undersøker en lignende problemstilling var testen fra PROBE-prosjektet. For det andre har jeg ikke funnet noen anerkjente eller utprøvde tester i kritisk tenkning som er offentlig tilgjengelige, verken for meg som testutvikler eller for lærere som ønsker å anvende slike tester i forbindelse med undervisning. Jeg valgte derfor å forsøke å utvikle et måleinstrument fra grunnen av, og hovedhensikten med masterprosjektet er selve testutviklingen og medfølgende utprøving, slik at den arbeidsintensive kvalitetssikringen av måleinstrumentet kan få nødvendig fokus. Jeg har likevel forsøkt å hente inspirasjon fra utprøvde variabler når det har vært mulig, og eventuelt rekontekstualisert disse slik at de passer med mine forskningsspørsmål. De fleste variablene i kartleggingsverktøyet har jeg utledet selv, ut fra den teoretiske forankringen som er gjort rede for i forrige kapittel. Hoveddelen av masteroppgaven vil derfor i stor grad omfatte redegjørelser av hvilke hensyn

som har blitt tatt ved utformingen av variablene og hvordan de ulike studiene har bidratt i utviklingen av det endelige kartleggingsverktøyet, med hensyn til kvalitetssikring.

### 3.1 Operasjonalisering og utvikling av kartleggingsverktøyet

Som beskrevet i kapittel 1.3, består kartleggingsverktøyet av tre deler. Hovedfokus i denne masteroppgaven er å utvikle en test som måler de ferdighetene som er viktige for å kunne tenke kritisk om naturfagsrelaterte temaer og tekster. Testutviklingen tilegnes derfor mest plass i beskrivelsen av studiets metode. I dette delkapittelet beskrives operasjonaliseringen av begrepet kritisk tenkning i kartleggingsverktøyet, med spesielt fokus på ferdighetstesten. Etersom ferdighetstesten er mest sentral i prosjektet, vil utviklingen av de to spørreskjemaene i kartleggingsverktøyet tilegnes mindre plass i masteroppgaven enn testutviklingen, som vist i Figur 3-1. Under gjøres det rede for utformingen av spørreskjema med bakgrunnsvariabler, ferdighetstesten og hvordan holdninger tilknyttet kritisk tenkning og naturfag ble operasjonalisert i et spørreskjema.

Figur 3-1: Ferdighetstesten er hovedfokus i operasjonaliseringen og kvalitetssikringen av kartleggingsverktøyet.



#### 3.1.1 Spørreskjema om bakgrunnsinformasjon

Hva som påvirker elevenes evne til å tenke kritisk er et interessant spørsmål, og undersøkes gjennom forskningsspørsmål 2.b. Derfor utformet jeg et spørreskjema til kartleggingsverktøyet som undersøker variabler som kjønn, skoletrinn, minoritetspråklig bakgrunn, lesning, karakterer i naturfag og norsk. Hvis variablene korrelerer med elevenes score på ferdighetstesten, indikerer dette at testen kan skille mellom ulike grupper av elever. Variablene ble samlet inn ved å stille generelle spørsmål i starten av undersøkelsen. Spørsmålene skulle i tillegg virke enkle og ufarlige for elevene slik at de ikke opplevde

undersøkelsen som avskrekkende. Konkrete begreper som kjønn og skoletrinn ble operasjonalisert gjennom spørsmål som «*hvilket skoletrinn tilhører du?*» med svaralternativene «*9. trinn*» og «*10. trinn*». Andre begreper som leseferdigheter var mer utfordrende å operasjonalisere, og har lavere umiddelbar validitet. Et eksempel på hvordan slike variabler ble operasjonalisert var at lesing ble undersøkt med to variabler. En variabel undersøker hvor mye elevene leser på fritiden, mens en annen undersøker hva elevene forsto forrige gang de leste i en naturfaglig lærebok. Det ble forsøkt å ta hensyn til klarhet i spørsmålsformuleringer og å begrense tolkningsrommet, for å oppnå reliable resultater (Patel et al., 1995; Ringdal, 2018).

### 3.1.2 Test av ferdigheter

Det er viktig å ha en klar oversikt over *innholdet* man ønsker å undersøke og hvordan dette skal testes, når man skal utvikle en test (Cohen et al., 2007). Som vist til i teorikapittelet, skal testen ta utgangspunkt konkretiseringen av begrepet kritisk tenkning fra Facione (1990a), og de ferdighetene som presenteres i studien. Dette delkapittelet vil ta for seg hvordan ferdighetene fra Facione (1990a) har blitt rekontekstualisert for naturfag på ungdomstrinnet, og operasjonalisert i en test, med fokus på å ta velbegrunnede avgjørelser og å kunne utføre kildekritikk av naturfaglig informasjon. I denne prosessen har jeg støttet meg til annen relevant teori om sammenhengen mellom kritisk tenkning og kildekritikk, argumentasjon og scientific literacy.

Testens *formål* er å utvikle et formativt kartleggingsverktøy. Det skal undersøke respondentens ferdigheter og holdninger, med hensikt om å være et verktøy lærere kan anvende for å gi tilbakemeldinger og revidere egen undervisning. Diagnostiske tester er ofte kriteriebaserte, men bør inneholde oppgaver med ulik vanskelighetsgrad for å finne ut hva eleven kan og ikke kan om kunnskapsområdet (Cohen et al., 2007). Testens *spesifikasjoner* handler om hvordan innholdet skal måles og vektlegges i testen. Hvis alle kunnskapsområdene i testen er like viktige bør mengden oppgaver og poeng være velbalansert mellom kunnskapsområdene (Cohen et al., 2007). Det bør også på forhånd fastsettes den omtrentlige størrelsen på testen, altså antallet og omfanget av oppgaver. For at hele kartleggingsverktøyet skal være gjennomførbart i løpet av en skoletime, inkludert spørreskjema og måling av holdninger, valgte jeg å holde hele kartleggingsverktøyet innenfor rammen av 50 spørsmål og oppgaver til sammen. Under testutviklingen kan det være

hensiktsmessig å anvende flere oppgaver enn dette, slik at oppgaver som ikke fungerer kan ekskluderes gjennom item-analyse.

For å operasjonalisere begrepet ferdigheter for kritisk tenkning i testen konstruerte jeg sammensatte variabler som videre vil omtales som kategorier og delferdigheter. Facione (1990a) identifiserte seks ferdigheter og seksten delferdigheter som viktige for kritisk tenkning, på et collegenivå (Tabell 2-2). Jeg anvender færre kategorier i ferdighetstesten enn antallet ferdigheter som Facione (1990a) identifiserte, slik at hver sammensatte variabel består av flere oppgaver. Dette gjøres fordi Cohen et al. (2007) anbefaler å anvende flere oppgaver for å undersøke hver kategori i testen, fordi den interne konsistensen av målingene i hver kategori blir høyere jo flere oppgaver som anvendes. I Tabell 3-1 beskrives de fire kategoriene som ble anvendt i testutviklingen, og testkategoriene er sammensatt av ulike ferdigheter og delferdigheter. Testkategoriene ble valgt fordi de til sammen sammenfattet elementer fra alle de originale ferdighetene. Kategoriene ble videre delt inn i delferdigheter, for å konkretisere de abstrakte kategoriene til mindre kunnskapselementer som er enklere å observere (Cohen et al., 2007).<sup>3</sup>

---

<sup>3</sup> Vedlegg 4.1 viser hvilke oppgaver som er tilknyttet hvilke delferdigheter i den fjerde utgaven av ferdighetstesten.

Tabell 3-1: Kategorier og delferdigheter som måles i testen

Kategori i testen	Beskrivelse	Delferdigheter
<b>Strategier for kritisk tenkning (S)</b>	Ferdigheter knyttet til å ha og bruke strategier for å kunne ta en avgjørelse og utøve kildekritikk. Dette inkluderer det å kunne gjøre rede for valg man har tatt og strategier man har brukt. Kategorien har tatt utgangspunkt i ferdighetene «Forklare» og «Selvregulering», og delferdigheten «vurdering av troverdighet»	<b>Formulere strategier (S1)</b> <ul style="list-style-type: none"> <li>• Kan begrunne og reflektere over egne strategier for hvordan man selv tar en gjennomtenkt konklusjon.</li> <li>• Kan forklare hvilke strategier man bruker for å være kildekritisk.</li> </ul>
		<b>Vurdere troverdighet (S2)</b> <ul style="list-style-type: none"> <li>• Kan vurdere en kildes troverdighet gjennom å stille enkle spørsmål om en tekst.</li> </ul>
<b>Forståelse av naturfaglig informasjon (N)</b>	Ferdigheter knyttet til å kunne forstå naturfaglig informasjon i ulike former, og å kunne forstå hvilken rolle informasjon er ment å ha i en tekst. Dette krever at eleven avkoder informasjonen, og forstår hvilken kontekst informasjonen er uttrykt. Eleven må også kunne identifisere argumenter i tekster, og forstå hvilken konklusjon de er ment å støtte. Kategorien har tatt utgangspunkt i ferdighetene «tolke» og «analysere»	<b>Avkode og tydeliggjøre informasjon fra tekster (N1)</b> <ul style="list-style-type: none"> <li>• Kan vise forståelse for naturfagsrelevant informasjon fra nettbaserte tekster.</li> <li>• Kan identifisere argumenter, påstander og begrunnelser i nettbaserte tekster.</li> </ul>
		<b>Avkode informasjon fra tabeller og illustrasjoner (N2)</b> <ul style="list-style-type: none"> <li>• Kan lese, avkode og analysere informasjon ut fra tabeller og illustrasjoner.</li> </ul>
<b>Vurdere ut fra kriterier<sup>4</sup> (A)</b>	Ferdigheten til å kunne vurdere argumenter og påstander ut fra gitte kriterier som logisk gyldighet og relevans. Kategorien har tatt utgangspunkt i delferdigheten «Vurdering av argumenter».	<b>Vurdering av argumenter (A1)</b> <ul style="list-style-type: none"> <li>• Kan vurdere argumenter ut fra kriterier som blir gitt om logisk gyldighet og om argumentet er relevant for problemstillingen.</li> </ul>
<b>Trekke velbegrunnede konklusjoner (K)</b>	Ferdigheten til å kunne identifisere nødvendig informasjon for å trekke troverdige konklusjoner, og bruke relevant informasjon for å trekke gjennomtenkte slutninger. Dette krever at man forstår de logiske konsekvensene som kan utledes fra informasjon.  Kategorien har tatt utgangspunkt i ferdigheten «Å trekke konklusjoner».	<b>Undersøke og vurdere bevis (K1)</b> <ul style="list-style-type: none"> <li>• Kan vurdere hvilke forklaringer som er mulige med grunnlag i bevis.</li> </ul>
		<b>Vurdere bevisgrunnlaget (K2)</b> <ul style="list-style-type: none"> <li>• Kan formulere egne alternative forklaringer til problemstillinger.</li> <li>• Kan vurdere om det er tilstrekkelig med informasjon for å ta en avgjørelse.</li> </ul>
		<b>Konkludere (K3)</b> <ul style="list-style-type: none"> <li>• Kan trekke nøye gjennomtenkte konklusjoner, med grunnlag i bevis.</li> </ul>

Testkategoriene gjorde begrepet mer konkret og målbart, og ble brukt for å operasjonalisere ferdighetene til det første utkastet av testen. Første utkast av testen ble brukt i studie 1, og var

<sup>4</sup> Testkategorien «Vurdere ut fra kriterier» og de oppgavene som var tilknyttet kategorien ble ekskludert fra testen etter studie 2. Dette beskrives nærmere i kapittel 4.2.



utgangspunktet for videre revidering. Første versjon av testen var delt inn i tre deler, med ulike temaer (Se vedlegg 1.2 og 1.3).

Del 1 var direkte inspirert av eksempeloppgaver fra andre tester av kritisk tenkning<sup>5</sup>. Her skulle elevene vurdere 14 argumenter ut fra kjennetegn på sterke og svake argumenter, som ble gitt som en stimulus før oppgavene. Som gjort rede for i kapittel 2 er det viktig at et sterkt argument er logisk gyldig og relevant for problemstillingen. Kjennetegnene ble tilpasset aldersgruppen, ved å fokusere på relevans til problemstillingen og at argumentene må bygge på faktabaserte premisser, og ikke meninger eller irrelevant informasjon. Jeg vurderte begrepet «logisk» som for abstrakt og ukjent for aldersgruppen, men forsøkte å utforme kjennetegnene og argumentene slik at elevene måtte anvende logikk for å vurdere om argumentene var sterke eller svake. Kjennetegnene ble presentert slik:

*Et **sterkt** argument kjennetegnes som:*

- *Direkte relevant til spørsmålet*
- *Bruker viktig og konkret informasjon*

*Et **svakt** argument kjennetegnes ved at det oppfyller et eller flere av disse kriteriene:*

- *Ikke direkte relevant for å svare på spørsmålet*
- *Uttrykker en mening, og ikke fakta.*
- *Informasjonen er lite konkret eller ikke viktig.*

Argumentene var tilknyttet to ulike «cases» hvor elevene ble presentert for naturfaglig informasjon. Case 1 omhandlet solcellepaneler, som vist til i eksempeloppgave 2, og case 2 handlet om genteknologi. Jeg forsøkte å basere enkelte av argumentene på kjente fallasier som autoritetsargumenter, popularitetsargumenter og skråplansargumenter. Elevene skulle ta utgangspunkt i at informasjonen i argumentene var sann, og kun vurdere om argumentet passet til kjennetegnene for sterke eller svake argumenter. Det ble anvendt lukkede oppgaver med to svaralternativer, og kun én nøkkel (AC). I tillegg ble det brukt en åpen oppgave om

---

<sup>5</sup> Det finnes en rekke gratis tester som baserer seg på de anerkjente testene. Eksempel på typiske oppgaver hvor man skal vurdere argumenter er tilgjengelige på gratis testen til AssessmentDay. *Watson Glaser Critical Thinking Appraisal*. Mercury House. Hentet 11.10.20 fra <https://www.assessmentday.co.uk/watson-glaser-critical-thinking.htm>

hvordan elevene tenkte når de vurderte argumentene. Del 1 skulle måle delferdigheten «vurdering av argumenter» og den åpne oppgaven målte delferdigheten «strategier».

#### 1.1. Bør Ola kjøpe solcellepanel for å spare penger?

Ola har et stort hus, og han lurer på om han skal kjøpe solcellepanel til å ha på taket. Et solcellepanel er en energikilde som samler solenergi fra sola, og gjør den om til elektrisk energi som Ola kan bruke i huset. Da trenger ikke Ola å kjøpe strøm fra strømselskapene.

Vurder om argumentene under er svake eller sterke, og velg alternativet som passer best.

##### 1.1 a) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Nei, han vil ikke spare penger på å kjøpe solcellepanel, fordi et solcellepanel er dyrt og har lav gjennomsnittlig levetid. Den går altså ofte i stykker før man har tjent inn det solcellepanelene kostet.

Sterkt argument

Svakt argument

*Eksempeloppgave 2: Argumentsoppgave fra del 1 (hentet fra testen i studie 2, men innholdet er likt førsteutkastet)*

I del 2 ble elevene presentert for et mysterium i form av en case. Denne delen hadde som hensikt å måle flere av delferdighetene kategorien «å trekke velgrunnede konklusjoner». Casen tok utgangspunkt i et virkelig mysterium om hvorfor det ble mindre tare på havbunnen ved Alaska (Estes et al., 1998). Målet med oppgaven var at elevene skulle vurdere hva som var den mest sannsynlige forklaringen på dette autentiske mysteriet. Oppgavene var av typen CDIS, som betyr at det var sammenheng mellom oppgavene, og informasjonen som ble gitt gjennom stimuluser, slik at det var mulig å følge elevenes resonnering gjennom oppgavesettet. Elevene ble gradvis presentert for flere mulige forklaringer og informasjonselementer gjennom oppgavene. Informasjonselementene omhandlet generelle fakta om økosystemer og informasjon om ulike faktorer som kan ha påvirket populasjonen av tang og tare. Ut fra informasjonselementene skulle elevene vurdere hvilke forklaringer som var mulige, og hvilke som kunne utelukkes. Selv om problemstillingen er autentisk, var noen av informasjonselementene konstruert for å kunne utelukke de andre forklaringene.

Oppgavene i del 2 skulle måle elevenes evne til å trekke velbegrunnede konklusjoner gjennom å undersøke og vurdere bevis, og Eksempeloppgave 3 og Eksempeloppgave 4 er eksempler på hvordan slike oppgaver ble formulert. Elevene ble spurt gjennom oppgavesettet om de hadde nok informasjon til å trekke en sikker konklusjon. Det ble i tillegg brukt kortsvarsoppgaver hvor elevene skulle formulere egne alternative forklaringer og forklare

hvordan de hadde tenkt, altså hvilke strategier de anvendte. Mengden informasjonselementer var svært omfattende, men ble introdusert gradvis gjennom oppgavesettet. På den måten var målet å undersøke når elevene ikke klarte å forholde seg til informasjonsmengden lenger, med utgangspunkt i når de begynte å svare feil på oppgavene.

#### 2.1 Bruk informasjonen, og vurder forklaringen:

- Tang og tare er sjøplanter som lever på bunnen av havet.
- På havbunnen utenfor Alaska har det gjennom flere år blitt observert at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt grunnet global oppvarming.

#### Konklusjon:

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

Konklusjon: Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur. \*

2.1 a) Er du enig i denne konklusjonen kan stemme?

Ja

Nei

*Eksempeloppgave 3: Konklusjonsoppgave fra del 2 (hentet fra testen i studie 2, men innholdet er likt førsteutkastet)*

2.3 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

*Eksempeloppgave 4: Konklusjonsoppgave fra del 2 (hentet fra testen i studie 2, men innholdet er likt førsteutkastet)*

I del 3 skulle elevene lese to autentiske tekster fra internett, og svare på en rekke oppgaver knyttet til tekstens troverdighet og innhold. Begge tekstene omhandlet kosthold, med fokus på kreftfaren ved å spise kjøtt. Den første teksten var valgt ut fordi den ga svært lite troverdig informasjon. Den var skrevet av bloggeren Sophie Elise på bloggen hennes, og var basert på informasjon fra en Netflix-produsert dokumentar og at hun «hadde søkt på google» (Isachsen, 2017). Teksten inneholder mange påstander som bryter med den allmenne oppfatningen om kosthold. Det påstås i blogginnlegget at å spise kjøtt er like kreftfremkallende som å røyke, ost sammenlignes med heroin og det hevdes at det er spor av dyreavføring i 90% av kjøtt som

selges i USA og ca. 45% av det som selges i Norge. I tillegg inneholder teksten mange skrivefeil. Selv om det ikke forventes at elevene kjenner til Sophie Elise fra før, er det flere faresignaler som tilsier at informasjonen i teksten er lite troverdig. Den andre teksten er skrevet av journalister i faktisk.no, og er et svar på blogginnlegget til Sophie Elise sine påstander om kjøtt (Akerbæk et al., 2017). I artikkelen refereres det til fagpersoner, forskning og anerkjente vitenskapelige institusjoner som WHO og kreftregisteret. Den nyanserer kreftfaren ved å spise kjøtt, gjennom å påpeke at noen kjøttprodukter er bevist kreftfremkallende, som rødt og bearbeidet kjøtt. Kreftrisikoen er likevel mye mindre enn ved røyking. Selv om elevene ikke kjenner til kreftregisteret eller WHO fra før, legges informasjonen frem på en oversiktlig måte, og nettsiden fyller flere kjennetegn for en troverdig kilde.

I den første utgaven av testen var flertallet av oppgavene i del 3 åpne. Åpne oppgaver ble anvendt for å få innsikt i hvordan elevene tenkte rundt tekstene, og for å forstå hva som kan forventes at elevene kan om kildekritikk. Noen oppgaver etterspurte om elevene kunne gjengi påstandene og begrunnelsene i tekstene. Hensikten med disse oppgavene var å måle elevenes forståelse av informasjonen i tekstene, og hvilken rolle informasjonen spilte i forfatterens argumentasjon. Det ble brukt lukkede oppgaver for å undersøke om elevene synes kildene var troverdige, og hvilken kilde som de mente var mest troverdig. Slike lukkede oppgaver ble etterfulgt av åpne oppgaver hvor elevene skulle begrunne svarene sine. I tillegg til oppgaver om de to kildenes troverdighet og innhold, blir elevene bedt om å oppgi faktorer de mener er viktige når man skal vurdere troverdighet til en nettside.

### 3.1.3 Spørreskjema om holdninger

I denne delen av oppgaven vil det redegjøres for hvordan holdningene ble operasjonalisert i spørreskjemaet. Holdningene ble målt for å undersøke om det er en korrelasjon mellom score på ferdighetstesten og positive holdninger til kritisk tenkning. Jeg har tatt utgangspunkt i de holdningskategoriene som ble presentert i teorikapitlet, med grunnlag i det teoretiske rammeverket fra Facione (1990a; 2000; 2015). *Holdninger* er ifølge Ringdal (2018) psykologiske tilstander som kun kan måles indirekte. En holdning er respondentens negative, nøytrale eller positive vurdering av et *holdningsobjekt*. Det er vanlig å måle holdninger gjennom vurderingsspørsmål, og til slike anvendes ofte en *likert-skala*. Respondenten blir da presentert for påstander som han skal ta stilling til. Svaralternativene i en likert-skala gir ordinale verdier, og kan for eksempel være «*helt enig – litt enig – verken enig eller uenig –*

*litt uenig – helt uenig*». Jeg har valgt å anvende en likert-skala, fordi de generiske svaralternativene gjør ulike variabler lettere å både besvare og sammenligne variablene siden de har samme svaralternativer (Friborg, 2010). Dette fjerner muligheten for å utvikle tilpassede svaralternativer til hvert spørsmål, og øker dermed tolkningsrammen til enkelte variabler (Ringdal, 2018).

Det er utfordrende å utvikle valide og reliable variabler til spørreskjemaer, men mange av rådene om hvordan man utvikler gode oppgaver gjelder også for spørreskjemaer. For å måle holdninger anbefaler Ringdal (2018) å bruke klart definerte holdningsobjekter i spørsmålene. Spørsmålene som stilles bør ikke være flerdimensjonale (Patel et al., 1995; Ringdal, 2018). Det vil si at spørsmålene kun spør og undersøker ett holdningsobjekt, og ikke flere. Man bør også benytte flere spørsmål til å måle holdninger til komplekse objekter, som kritisk tenkning. Jeg har valgt å anvende fire spørsmål i hver holdningskategori, for å begrense omfanget av kartleggingsverktøyet, selv om dette går på bekostning av den interne konsistensen, og muligheten til å ekskludere spørsmål etter pilotering.

Patel et al. (1995) anbefaler at stammen i spørsmålet bør være så kort som mulig, og ikke inneholde overflødig informasjon og instruksjoner. Det er en av grunnene til at jeg har anvendt en likert-skala med generiske svaralternativer. Jeg forsøkte å unngå unødvendig generelle spørsmål, fordi dette kan føre til tvetydigheter (Patel et al., 1995). Flere av spørsmålene handler derfor spesifikt om skole og informasjon fra internett. Utfordringen med dette kan være at elever som har generelle positive holdninger til kritisk tenkning virker negativt innstilte, hvis de har negative holdninger til skole eller internett. Det er viktig at spørsmålet er enkelt å lese, tolke og forstå, og man bør bruke så enkle begreper som mulig, dersom dette ikke påvirker innholdet i spørsmålet (Cohen et al., 2007). Man bør ifølge Ringdal (2018) tilpasse språket til målgruppen man skal undersøke. I dette studiet er målgruppen elever på ungdomsskolen, hvor enkelte kan ha et svært begrenset ordforråd. Derfor har jeg forsøkt å bruke et enkelt språk, med fokus på holdningsobjekter som alle har et forhold til. Jeg unngikk formuleringer som kunne føre til at spørsmålene ble ledende, eller inneholdt verdiladede ord (Patel et al., 1995; Ringdal, 2018).

Det finnes utprøvde måleinstrumenter for holdninger til kritisk tenkning, som *California Critical thinking disposition inventory* (2000), som Facione har vært med å utvikle. Igjen er utfordringen at testen ikke er tilgjengelig uten lisens. Jeg valgte derfor å utvikle et eget

spørreskjema for å måle holdningene, med utgangspunkt i holdningskategoriene som anvendes i Facione et al. (2000). Dette ga samtidig muligheten til å tilpasse holdningsobjektene til skole og naturfag, og tilpasse språket til målgruppen. Det oppgis eksempler på holdningsspørsmål i Facione (2015) og Facione (2000). Jeg tilpasset og rekontekstualisert de relevante eksemplene fra Facione og anvendte de i mitt spørreskjema. I Tabell 3-2 er holdningskategoriene beskrevet med tilhørende variabler. For å begrense antall spørsmål i måleverktøyet har jeg valgt å slå sammen enkelte av de holdningskategoriene som Facione et al. (2000) har utformet. De kategoriene som ble sammenslått var relevante for hverandre, og det var ikke et klart skille mellom dem. Utfordringen med en slik sammenslåing er at holdningskategorien blir flerdimensjonal, og kan påvirke den interne konsistensen i kategorien.

Hver holdningskategori i spørreskjemaet består av fire spørsmål. Unntaket er holdningskategorien «*Holdninger til kritisk tenkning i skolen*», som består av tre spørsmål. Spørreskjemaet med holdninger ble først utprøvd i studie 3 i prosjektet. Det ble lagt til to selvutviklede holdningskategorier: «*Holdninger til naturfag*» og «*holdninger til kritisk tenkning i skolen*» for å undersøke om det er en korrelasjon mellom elevenes svar på de holdningsobjektene og testscore på ferdighetstesten.

Tabell 3-2: Holdningskategorier i kartleggingsverktøyet

Kategori i testen	Beskrivelse	Item nr.	Tilknyttede spørsmål (items)	Kilde / inspirert av
<b>Intellektuell nysgjerrighet</b> (Inquisitiveness)	Kategorien skal måle respondentens intellektuelle nysgjerrighet. (Facione et al., 2000)  Det motsatte av intellektuell nysgjerrighet er likegyldighet til ny informasjon (Facione, 2000)	I-1	Å studere nye ting hele livet ville vært fantastisk *.	Facione, 2000
		I-2	Ofte legger jeg lite innsats i å lære noe nytt på skolen, før jeg vurderes i det. (negativ)**	Facione, 2015
		I-3	Jeg bruker mye tid på å lese om ting jeg er nysgjerrig på.	Egenformulert
		I-4	Jeg kan bruke timevis på å lese om ukjente temaer på internett.	Egenformulert
<b>Åpenhet for ny informasjon</b> (Open-Mindedness)	Kategorien skal måle respondentens holdninger knyttet til åpenhet for ny informasjon og andres meninger, selv om dette motstrider egne verdier og meninger. (Facione et al., 2000)  Det motsatte av åpenhet er intoleranse for divergente meninger og motstridene informasjon (Facione, 2000)	A-1	Andre har rett på sine meninger, men jeg trenger ikke høre dem. (negativ) *	Facione, 2000
		A-2	Jeg kan gjøre narr av og erte andre for deres meninger og verdier. (negativ) *	Facione, 2015
		A-3	Jeg er åpen for ny informasjon, selv om jeg er ganske sikker i saken.	Egenformulert
		A-4	Jeg liker å lese på internett om meninger jeg er uenig med, for å få innsikt i hvordan andre tenker.	Egenformulert
<b>Selvtillit</b> (Critical Thinking Self-Confidence og Analyticity)	Kategorien er formulert med utgangspunkt i to av kategoriene som Facione et al. (2000) referer til som « <i>Critical Thinking Self-Confidence</i> » og « <i>Analyticity</i> »  Kategorien kan oppsummeres som tiltro til egne evner til å ta gode avgjørelser og tenke kritisk.	Se-1	Jeg er stolt av hvor flink jeg er til å komme opp med kreative løsninger på problemer. *	Facione, 2000
		Se-2	Når en oppgave er komplisert får jeg en følelse av panikk. (negativ) *	Facione, 2000
		Se-3	Jeg gir meg ikke når oppgavene er vanskelige.	Egenformulert
		Se-4	Jeg er god til å vurdere om man bør stole på en nettside eller ikke.	Egenformulert
<b>Sannhetssøkende</b> (Truth-Seeking)	Kategorien skal måle om respondentenes ønske om å søke til sannheten, selv om dette kan medføre at respondenten selv må endre holdninger, verdier og meninger (Facione et al., 2000).  Det motsatte av å være sannhetssøkende er å være uærlig i sin bruk av informasjon (Facione, 2000).	Sa-1	Det viktigste i en diskusjon er å vinne. (negativ) **	Facione, 2015
		Sa-2	Uansett hva som skjer bør man holde på verdiene sine, og ikke revurdere dem. (negativ)	Egenformulert
		Sa-3	Det er viktig for meg at den kunnskapen jeg har er så sikker som mulig.	Egenformulert
		Sa-4	Jeg velger ut informasjon som passer med det jeg mener. (negativ) **	Facione 2015

<b>Betenksomhet</b> ovenfor avgjørelser  (Systematicity og Maturity)	Kategorien er formulert med utgangspunkt i to av kategoriene som Facione et al. (2000) referer til som «Systematicity» og «Maturity». Kategorien kan oppsummeres som hvilke holdninger respondenten har til hvor velbegrunnede avgjørelser bør være.	B-1	Når jeg skal ta en avgjørelse, forsøker jeg først å finne mest mulig informasjon om saken. *	Facione, 2000
		B-2	I stedet for å jobbe grundig med å løse en oppgave, spør jeg en klassekamerat om svaret. (negativ) **	Facione, 2015
		B-3	Jeg tar beslutninger raskt, og tenker på konsekvensene først når de kommer. (negativ)	Egenformulert
		B-4	Jeg sjekker alltid med flere kilder når jeg leser om et tema jeg kan lite om fra før.	Egenformulert
Holdninger til <b>naturfag</b>	Denne kategorien undersøker respondentens holdninger til læringsutbyttet fra naturfag.	N-1	Jeg er nysgjerrig på det vi lærer i naturfag.	Egenformulert
		N-2	Jeg synes det man lærer i naturfag er viktig å kunne når man skal delta i demokratiske valg.	Egenformulert
		N-3	På fritiden tenker jeg over ting jeg lærer i naturfag.	Egenformulert
		N-4	Naturfagskunnskapene jeg har hjelper meg med å være kildekritisk.	Egenformulert
Holdninger til <b>kritisk tenkning</b> i skolen	Kategorien er en samling av spørsmål om viktigheten og læringsutbytte om kritisk tenkning som skolen har gitt eleven	H-1	Å lære å tenke kritisk er det noe av det viktigste man lærer på skolen.	Egenformulert
		H-2	Jeg har lært mye hvordan man skal tenke kritisk på skolen.	Egenformulert
		H-3	På skolen har jeg lært hvordan man kan være kildekritisk.	Egenformulert

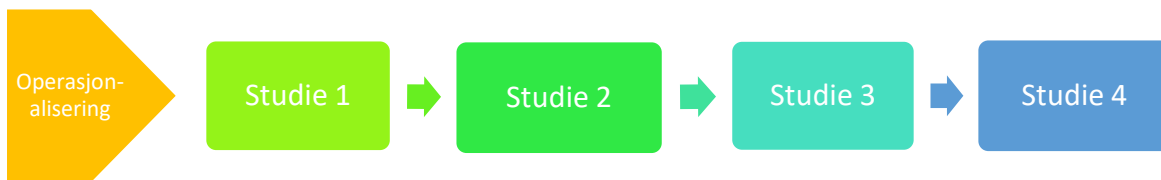
\*Oversatt av Sebastian, med få endringer

\*\*Rekontekstualisert for tilpasning til elevgruppen og tema

### 3.2 Studiets oppbygging og utvalg

I denne delen av oppgaven redegjøres det for hvordan prosjektet er bygget opp, og for de ulike utvalgene. Masterprosjektet ble gjennomført suksessivt i fire studier, som vist i Figur 3-2, hvor de tre første studiene bidro til å kvalitetssikre kartleggingsverktøyet på ulike måter, før kartleggingsverktøyet ble prøvd ut på et større utvalg i studie 4. Fordelen med et slikt forskningsdesign er at det var mulig å revidere måleinstrumentet mellom studiene, med utgangspunkt i resultatene.





Figur 3-2: Prosjektets suksessive oppbygning

En utfordring med kvantitative måleinstrumenter er at det ikke er mulig å endre på spørsmålene og oppgavene når datainnsamlingen har begynt. Det er derfor viktig å kvalitetssikre variablene gjennom en *pilottest* før undersøkelsen. En pilottest har som hensikt å bidra til å øke instrumentets reliabilitet, validitet og praktiske anvendelse (Cohen et al., 2007). Å pilotere en test vil si at man gjør en mindre utprøving av måleinstrumentet (Cohen et al., 2007), og dette kan også gjøres med spørreskjemaer (Friborg, 2010). Tradisjonelt innen testutvikling er det vanlig å utforme mange oppgaver, som man pilottester, og ekskluderer de oppgavene som bidrar negativt til testens kvalitet, etter å ha gjennomgått en item analyse (Cohen et al., 2007). For at det skal være enkelt å ekskludere oppgaver fra en tradisjonell pilottest er det formålstjenlig å anvende mange separate testledd. Denne metoden egner seg ikke i mitt prosjekt, ettersom ferdighetstesten består av mange CDIS-oppgaver, altså oppgaver som er knyttet opp mot hverandre og stimuluser. Derfor har jeg valgt å pilotere og revidere ferdighetstesten gjennom flere studier, slik at endringene kan utprøves. Dette er i tråd med Cohen et al. (2007) anbefaling om å pilotere en test flere ganger, for å undersøke om endringene fikk ønsket resultat. Man kan skille mellom det som kalles prepilot, hvor testleddene utvikles, og pilot, hvor testens kvalitet undersøkes (Cohen et al., 2007). De første tre studiene i testutviklingen anses som mindre prepiloter, ettersom studie 1, 2 og 3 bidro til å utvikle kartleggingsverktøyet slik det ble anvendt i det fjerde studiet. Det fjerde studiet er en pilotstudie hvor testens kvalitet undersøkes, men hvor samtidig elevenes resultat på kartleggingsverktøyet undersøkes.

Dette masterprosjektet operer med fire utvalg, slik at versjoner av kartleggingsverktøyet stegvis kan utprøves, revideres og kvalitetssikres. Under beskrives den generelle utvalgsstrategien og det redegjøres for avveininger som har blitt gjort. Videre beskrives de fire studiene av prosjektet og tilhørende utvalg. Etske hensyn og avveininger i forhold til personvern behandles senere i kapittelet.

I denne masterstudien er det anvendt *ikke-sannsynlighetsutvalg*, og vil si at respondentene ikke har lik sjanse for å trekkes ut til en undersøkelse som resten av populasjonen (Ringdal,

2018). Fordelen med å anvende ikke-sannsynlighetsutvalg er at det er enklere å gjennomføre enn et sannsynlighetsutvalg, men skjer på bekostning av at man bør være mer forsiktig med å generalisere resultatene for hele populasjonen (Aidley, 2018). Skolen som utvalget tilhører er og har vært min arbeidsplass i fem år. Et utvalg hentet fra egen arbeidsplass kan kalles for et *bekvemmelighetsutvalg*. Et bekvemmelighetsutvalg er når utvalget blir valgt fordi det er enkelt å få til (Bjørndal & Hofoss, 2004). Fordelen med å anvende egen arbeidsplass er at man har tilgang til respondenter det kanskje ellers ikke ville vært mulig å rekruttere, og kan medføre at utvalget blir større. Samtidig fordrer det at man trår varsomt med hensyn til etiske avveininger som det redegjøres for i kapittel 3.4. Skolen ligger i Oslo, og har et stort mangfold av elever, både i etnisitet, kultur og sosioøkonomisk bakgrunn.

I prosjektet fokuseres det på ungdomsskoleelevers kritiske tenkning. Ungdomstrinnet ble valgt med tanke på allmenndannende rollen til naturfaget og kritisk tenkning for denne aldersgruppen. Grunnskolen har en særegenrolle i skolens allmenndannende mandat. Begrepet kritisk tenkning består av holdninger og ferdigheter som det er prekært at elevene har innehar i møte med fremtiden. Valget om å undersøke elever på 9. og 10. trinn ble gjort fordi elevene er eldre enn elever på 8. trinn, og det skjer mye modning i aldersgruppen. Som vist i Tabell 3-3 er alle utvalgene hentet fra 9. og 10. trinn, med unntak av studie 2. Gjennom å velge elever fra to ulike trinn var det mulig å sammenligne trinneses resultater på testen, og få en indikasjon på om elever utvikler ferdighetene som testen måler i løpet av et skoleår.

Tabell 3-3: Oversikt over studiene og utvalgene i masterprosjektet

Del av prosjektet	Studie 1	Studie 2	Studie 3	Studie 4
Utvalgsnummer	Utvalg 1	Utvalg 2	Utvalg 3	Utvalg 4
Antall (n)	15	8	11	76
Trinn / aldersgruppe	9. trinn	4 venner og 4 naturfagslærere	10. trinn	9. og 10. trinn
Analyseverktøy	Skjønnsmessig vurdering og tolkning	Revideringer etter tilbakemeldinger	Item-analyse gjennom SPSS (Vanskelighets-grad, diskrimineringsevne og Cronbach Alfa)	Item-analyse og korrelasjons-analyser gjennom SPSS

### 3.2.1 Studie 1

Det første utkastet av testen ble utprøvd på utvalg 1 i studie 1. Utvalg 1 er rekruttert fra en klasse fra 9. trinn som jeg selv underviser i naturfag, og undersøkelsen ble gjennomført i forbindelse med at vi arbeidet med temaet kritisk tenkning i naturfag. Versjonen av testen

som ble anvendt inneholdt de tre delene som ble beskrevet i delkapittel 3.1.2. Del 1 og 2 ble gjennomført først, og del 3 ble gjennomført tre uker senere som en separat test. Flere åpne oppgaver ble anvendt i denne versjonen av testen enn i det endelige kartleggingsverktøyet. Cohen et al. (2007) anbefaler å anvende flere åpne spørsmål i prepilotstudier, for å få innsikt i elevenes refleksjoner. Elevenes svar på studie 1 ble brukt som et utgangspunkt for å videreutvikle kategorier og oppgaver. Alle delene ble gjennomført på itsearning, men elevene fikk også utskrevne kopier av informasjonen fra testen, og det er denne som er vedlagt (Se vedlegg 1.2 og 1.3). Dette ble gjort slik at elevene ikke trengte å trykke seg gjennom den digitale testen for å finne informasjon fra tidligere oppgaver. Spesielt del 2 og 3 var avhengig av at elevene forholder seg til informasjon som har blitt gitt tidligere i testen. Det ble ikke gjennomført statistiske analyser i studie 1, med unntak av å undersøke oppgavens vanskelighetsgrad. Formålet med studie 1 var å:

- A. Undersøke hvilke oppgaver og hva slags oppgaveformuleringer som gir best innsikt i elevenes kompetanse. Hvilke begreper, oppgaveformuleringer og instruksjoner som elevene synes var vanskelige ble undersøkt gjennom observasjon og gjennom svarene deres i testen. Observasjoner ble notert i et observasjonsskjema (Se vedlegg 1.1)
- B. Undersøke elevenes refleksjoner og mulige feiloppfatninger som oppsto i møte med oppgavene.
- C. Undersøke gjennomføringstiden, og hvor omfattende ferdighetstesten er.
- D. Utarbeide et løsningsforslag til de resterende åpne oppgavene, med utgangspunkt i svarene til elevene. Cohen et al. (2007) presiserer at det er viktig å planlegge hvordan poengutdeling skal utregnes når en test skal utvikles.
- E. Få en indikasjon på hvilket ferdighetsnivå man kan forvente fra 9. trinns elever.

### 3.2.2 Studie 2

Mellom studie 1 og 2 ble testen revidert, og mange av de åpne oppgavene ble gjort om til lukkede oppgaver. Ferdighetstesten ble også tilpasset UiO nettskjema, slik at de videre undersøkelsene kunne gjennomføres anonymt. I tillegg ble et tidlig utkast av spørreskjemaene utformet, slik at kartleggingsverktøyet besto av mer enn bare ferdighetstesten. Utvalg 2 besto av en gruppe på fire venner og en gruppe på fire naturfagslærere. Begge gruppene gjennomførte undersøkelsen mens jeg var til stede, og svarte på spørsmål og ga tilbakemelding på kartleggingsverktøyet underveis. Gruppen med venner gjennomførte kartleggingsverktøyet først, og revideringer ble gjennomført før lærergruppen gjennomførte

den. Lærergruppen ble valgt fordi deres naturfagsklasser skulle delta i studie 3 og 4. På den måten fikk lærerne erfaringer med testen, og ble bedre forberedt til å teste klassene sine selv. Formålet med studie 2 var å:

- A. Undersøke og gi tilbakemeldinger om ordbruken i kartleggingsverktøyet, og gi tilbakemelding på hva er vanskelig å forstå generelt, og hva som kan være vanskelig å forstå for en 9. klassing. Å øke lesbarheten i en test er en av hovedgrunnene til å gjennomføre en prepilot (Cohen et al., 2007).
- B. Få tilbakemeldinger på oppgavene, instruksene og spørsmålene i kartleggingsverktøyet. Dette bør ifølge Cohen et al. (2007) gjøres for å teste oppgavene, øke validitet og reliabilitet, og eliminere uklarheter i testen. Det ble lagt spesielt vekt på hvordan oppgavene kan tolkes og om de lukkede oppgavene kun har ett svaralternativ som kan tolkes som nøkkelen. I tillegg ønsket jeg å få innsikt i hva slags refleksjoner som kan gjøres rundt oppgavene. Begge gruppene ble spurt om de kunne gi tilbakemeldinger på hvilke instruksjoner og informasjon i teksten som var nødvendig og hva som var overflødig.

### 3.2.3 Studie 3

Med utgangspunkt i tilbakemeldingene fra studie 2 ble kartleggingsverktøyet revidert, slik at kartleggingsverktøyet i studie 3 kunne være så likt det som ble utprøvd på utvalg 4, både i utforming og kvalitet. Undersøkelsen i studie 3 ble gjennomført anonymt på et utvalg på 11 elever i 10. klasse. Utvalget ble rekruttert gjennom naturfagslærer og det var frivillig å delta. Når studie 3 ble gjennomført var ungdomsskolene i Oslo satt til rødt nivå, grunnet covid-19-situasjonen. Dette førte til at klassen ble delt i to kohorter, og det var kun en av kohortene som gjennomførte studie 3. Både faglærer og jeg var til stede under undersøkelsen, og det var faglæreren som ga informasjon om studien, kartleggingsverktøyet og hva det vil si å delta. Elevenes anonyme resultater, i tillegg til observasjon og elevenes spørsmål underveis, utgjør datagrunnlaget fra studie 3. Resultatene fra studie 3 bør undersøkes med forsiktighet, ettersom utvalgsstørrelsen er for lite for å gi pålitelige resultater ved item-analyse. Formålet med studie 3 var å:

- A. Undersøke hvor godt oppgavene og spørsmålene i det reviderte kartleggingsverktøyet fungerte gjennom item-analyse. Det ble gjennomført analyser av oppgavenes diskriminering og vanskelighetsgrad, i tillegg til at både kategorier i testen og

holdningskategoriene ble testet med Cronbach Alfa. Oppgaver hvor det var mulig å identifisere mulig årsak til svake resultater på item-analyse ble revidert før studie 4. I tillegg ble løsningsforslaget på de åpne oppgavene utprøvd.

- B. Observere elever som gjennomfører testen, med fokus på gjennomføringstid, og å registrere hva elevene lurer på i løpet av undersøkelsen. Spørsmål om oppgaver og instruksjoner ble notert i et observasjonsskjema.
- C. Teste instruksjonene om undersøkelsen, som ble gitt av faglæreren. Faglæreren fikk på forhånd tilsendt en manual for hva slags informasjon som skulle gis til foresatte og elever før studie 3, og til elevene under undersøkelsen.
- D. Teste ut tekniske aspekter ved kartleggingsverktøyet på elever. Tekniske aspekter omhandler formatet og designet på måleinstrumentet, og hvordan dataen kodes for å gjennomføre statistiske analyser.

#### 3.2.4 Studie 4

I det fjerde og siste studiet ble kartleggingsverktøyet utprøvd på et større utvalg. Hvordan studie 4 ble gjennomført og hvordan datamaterialet ble analysert vil derfor presenteres mer inngående enn de andre studiene, ettersom det først er i studie 4 at det er mulig å besvare forskningsspørsmålene i dette masterprosjektet.

Studie 4 hadde to hovedformål. For det første skulle studie 4 være et ledd i testutviklingen, og anvendes som en siste kvalitetsundersøkelse og pilottest. De utvalgte kvalitetsmarkørene ble anvendt for å undersøke kvaliteten til oppgavene, delferdighetene, testkategoriene og testen. I henhold til Cohen et al. (2007) sin anbefaling om pilotering, er kartleggingsverktøyet som anvendes i studie 4 mer omfattende og inneholder flere variabler enn det som er ønskelig i det ferdige måleinstrumentet. Dette er gjort for å kunne ekskludere de oppgaver og holdningsvariabler som ikke bidrar positivt til validiteten eller reliabiliteten. Tradisjonelt vil en pilotstudie ha enda flere oppgaver, og slik at man kan ha en streng utvelgelse av oppgaver etter item-analysen. Den suksessive oppbygningen i dette prosjektet begrenser denne muligheten. Utvalg 4 besto av fire skoleklasser fra 9. og 10. trinn, og de gjennomførte kartleggingsverktøyet, som har blitt revidert gjennom studie 1, 2 og 3. For å gjennomføre pålitelige analyser av diskrimineringene, vanskelighetsgrad og reliabilitet i testen anbefaler Cohen et al. (2007) å inkludere minst 50-100 respondenter i utvalget, og størrelsen på utvalg 4 er derfor egnet for å gjennomføre en pilottest. For det andre ble data om elevene samlet inn i

studie 4, og dette datamaterialet skulle besvare studiens andre forskningsspørsmål. Neste delkapittel tar for seg hvordan datamaterialet fra studie 4 ble analysert.

Manualen fra studie 3 ble revidert og anvendt i studie 4. Faglærerne gjennomførte undersøkelsen selv i henhold til manualen, og fylte ut utdelt observasjonsskjema (se vedlegg 4.3), uten at jeg var til stede. Denne vurderingen ble tatt på grunn av covid-19-situasjonen, for å minske smittefaren. Jeg gjennomførte kartleggingsverktøyet i en av klassene selv, hvor jeg er faglærer. I denne klassen ble det gjennomført en klassesamtale om kartleggingsverktøyet etter undersøkelsen, for å få innsikt i hva elevene synes var utfordrende, hvordan de tenkte rundt utvalgte oppgaver og elevenes refleksjoner rundt oppgavene i testen.

### 3.3 Analyse

Dette delkapittelet tar for seg hvordan datamaterialet fra studie 4 ble klargjort, bearbeidet og analysert. Hensikten med dette masterprosjektet kan anses å være todelt. For det første skulle det utformes en reliabel og valid test som kan måle elevers ferdigheter innen kritisk tenkning i naturfag. Testutviklingen og kvalitetssikringen av ferdighetstesten fokuseres på i forskningsspørsmål 1. Analysemetodene for å undersøke kartleggingsverktøyets kvalitet er allerede beskrevet, men hvordan de vil anvendes i dette masterprosjektet vil beskrives kort i dette kapittelet. Den andre delen av masterprosjektet handler om å utnytte datamaterialet fra testen utprøving i studie 4 for å få innsikt i elevenes tenkning, delferdighetenes vanskelighetsgrad, og undersøke mulige korrelasjoner mellom ulike variabler og testscore. De analysemetodene som ble anvendt for å undersøke prosjektets andre forskningsspørsmål presenteres nederst i dette delkapittelet. Først gjøres det rede for hvordan datamaterialet ble bearbeidet før analysene.

#### 3.3.1 Bearbeiding og klargjøring av datamaterialet

Kvantitative data må bearbeides før de kan analyseres, og denne delen av oppgaven tar for seg hvordan dette ble gjort i studie 4. Den versjonen av kartleggingsverktøyet som ble utprøvd på utvalg 4 var et resultat av studie 1, 2 og 3, og den kvalitetssikring og revidering som allerede var gjennomført på måleinstrumentet. Analyseverktøyet som anvendes er Statistical Package for the Social Science (SPSS). Datamaterialet må omkodes og klargjøres før det kan bearbeides i SPSS (Eikemo & Clausen, 2012). I mitt prosjekt handler omkoding om endring av skriftlige svaralternativer til tallverdier. Dette ble gjort i både studie 3 og 4.

Datarensgering kan gjøres ved å undersøke resultatene for åpenbare feil (Pallant, 2016). For å finne slike feil forsøkte jeg å undersøke om svarene til respondentene samsvarte med svarskalaen for variabelen. Dette ble gjort før og etter at nominale og ordinale verdier er omkodet, for å unngå at det skjedde feil i omkodingen. Et eksempel er variabelen kjønn, med svaralternativene «gutt» og «jente», som omkodes til verdiene «1» og «2». I studie 4 ble det i tillegg gjort en skjønnsmessig vurdering av besvarelsene til respondentene, slik at besvarelser til useriøse respondenter kunne ekskluderes fra datamaterialet. Useriøse besvarelser kan påvirke studiens reliabilitet negativt. Jeg så gjennom datamaterialet, og undersøkte om det var besvarelser hvor det var tydelig at respondenten bare hadde trykket på tilfeldige svar, og om det var useriøse svar på de åpne oppgavene. I tillegg undersøkte jeg besvarelsene til de elevene som ble ferdige med kartleggingsverktøyet raskere enn det som burde være praktisk gjennomførbart, hvis de tok testen seriøst.

### 3.3.2 Kvalitetssikring av testen

En viktig del av testutviklingen er å kvalitetssikre måleinstrumentet. Hvordan ferdighetstesten ble kvalitetssikret er gjort rede for i tidligere delkapitler, og hva som ble gjort for å kvalitetssikre testen i studie 4 vil kun beskrives kortfattet her. I studie 4 tok item-analysen utgangspunkt i oppgavens diskrimineringssevne, vanskelighetsgrad og hvordan oppgavene påvirker den interne konsistensen i delferdigheten og kategorien de tilhører. Anbefalingene jeg fulgte angående vanskelighetsgrad ( $p$ ) var at oppgavene bør ha en  $p$ -verdi mellom 0,85 eller 0,15. I tillegg var det ønskelig å ha en gjennomsnittlig  $p$ -verdi på ca 0,5 på testen, til tross for at testen er kriteriebasert. Diskrimineringssevnen ( $d$ ) til hver oppgave bør være så høy som mulig. Oppgavene bør ha høyere  $d$ -verdi enn 0,3, for å ikke påvirke testens konsistens negativt (Lie et al., 2005). De oppgavene som ikke hadde ønskelig vanskelighetsgrad eller diskriminerte godt nok ble foreslått endret, erstattes eller ekskluderes fra testen. Valget om man bør ekskludere oppgaver fra testen, gjøre endringer eller bevare oppgaver i sin helhet er ikke hugget i sten. Dersom man skal gjøre endringer, bør man ta hensyn til innholdsvaliditeten i testen (Cohen et al., 2007). I mitt prosjekt betyr at jeg foreslo å beholde utvalgte oppgaver i testen til tross for at de for eksempel har lav vanskelighetsgrad eller svak diskrimineringssevne, hvis fjerning av oppgavene ville ført til tap av sentrale aspekter ved det teoretiske begrepet kritisk tenkning.

For å undersøke om hvor reliable resultatene på testen er anvendes kvalitetsmarkøren Cronbach alfa, som er et mål på intern konsistens. Siden testen måler ulike ferdigheter og

delferdigheter, ble hver enkelt kategori i testen analysert med Cronbach alfa. Dette ble også gjennomført på spørreskjemaet for holdninger i kartleggingsverktøyet. Som nevnt i tidligere delkapittel er Cronbach alfa påvirket av hvor mange variabler som undersøkes sammen. I SPSS får man opp hvor mye hver variabel påvirker korrelasjonen, og hvor sterk den interne konsistensen ville vært hvis man fjernet en av variablene. Dette var et utgangspunkt for å fjerne variabler som bidro negativt til den interne konsistensen i en delferdighet eller kategori. I kategorier og delferdigheter hvor det ble brukt ulikt antall distraktorer eller oppgavene ga ulik uttelling av poeng ble standardisert Cronbach alfa anvendt. Testens interne konsistens som helhet ble også undersøkt. I tillegg ble testens standardfeil undersøkt, for å få et mål på den forventede feilmarginen i testen, og fordi Cronbach og Shavelson (2004) mener at standardfeil er det viktigste målet på reliabiliteten til et måleinstrument. Sensorreliabiliteten ble tatt hensyn til gjennom at testen i stor grad besto av lukkede oppgaver, og ved å utvikle konkrete vurderingskriterier for de tre åpne oppgavene i testen.

Det ble kun gjennomført en skjønnsmessig vurdering av hvor godt ivaretatt validiteten til måleinstrumentet i studie 4, gjennom å undersøke samsvaret mellom de teoretiske begrepene og måleinstrumentet. Umiddelbar validitet, begrepsvaliditet og innholdsvaliditet ble undersøkt gjennom en slik vurdering. En av fordelene med en suksessiv testutvikling er at oppgavene kunne revideres med utgangspunkt i elevenes besvarelser, samtidig som jeg kunne vurdere alle endringer opp mot de teoretiske begrepene. Suksessiv testutvikling er et omfattende arbeid, og på grunn av prosjektets omfang er det ikke gjennomført kvantitative validitetsanalyser som for eksempel faktoranalyse eller korrelasjonsanalyse opp mot et kjent måleinstrument. En slik systematisk analyse av validiteten kan gi et mer håndfast mål, og bør prioriteres dersom kartleggingsverktøyet skal videreutvikles etter denne studien. Det vil si at testens reliabilitet har blitt mer undersøkt enn validiteten i dette masterprosjektet. Årsaken til dette er at et måleinstrumentets reliabilitet setter en øvre grense for hvor valide resultatene kan bli (Patel et al., 1995). Konsekvensen av dette er at testens reliabilitet kan anslås mer nøyaktig enn validiteten ved masterprosjektets ende.

Det ble også testet for normalfordeling i score på ferdighetstesten. Dette ble gjort gjennom SPSS, og resultatet på en Kolmogorov-Smirnov test ble tolket. Hvis signifikansverdien på en slik test er over 0,05 indikerer dette at det er en normalfordeling i utvalget (Pallant, 2016). Det er også anbefalt å undersøke fordelingen av resultater ved å studere et histogram med



respondentenes testscore (Pallant, 2016), og dette ble gjort som et supplement til resultatene fra Kolmogorov-Smirnov test.

### 3.3.3 Analyser av testens resultater

I denne delen av oppgaven redegjøres det for hvilke analyser som ble gjort for å undersøke hva kartleggingsverktøyet kan fortelle om utvalget. I hovedsak handlet dette om å undersøke korrelasjoner mellom score på ferdighetstesten og variabler som kjønn, klassetrinn, karakterer, selvrapportert leseferdighet og holdninger til kritisk tenkning. Under presenteres de kvantitative testene som har blitt anvendt for å analysere datamaterialet.

Det er vanlig i kvantitative studier å presentere datamaterialet på en oversiktlig måte. En slik oversikt kan deskriptive analyser gi, og da utvikles det beskrivende statistikk. Hensikten med dette er å beskrive det innsamlede materialet ved hjelp av tall (Patel et al., 1995). Man får et godt bilde av fordelingen i datamaterialet hvis man kjenner gjennomsnittet og standardavviket (Patel et al., 1995), og derfor oppgis dette i resultatene fra studie 4. *Standardavvik* er det mest kjente spredningsmålet (Patel et al., 1995). Standardavvik kan regnes ut på enkeltspørsmål og på hele testen, og er et mål på gjennomsnittlig avstand fra gjennomsnittet (McAlpine, 2002). *Gjennomsnitt* er det mest anvendte sentralmålet, og gir innblikk i den gjennomsnittlige tendensen i utvalget (Ringdal, 2018).

For å undersøke mulige sammenhenger mellom variabler for utvalget og testscore ble det anvendt uavhengige T-tester, Variansanalyse (ANOVA), Pearsons korrelasjonstest og Spearman's rang-korrelasjonstest. Testene egner seg for å undersøke korrelasjoner eller undersøke forskjeller i gjennomsnittscore mellom elevgrupper (Ringdal, 2018).

Uavhengige T-tester kan anvendes for å sammenligne gjennomsnittsscore for to ulike grupper (Ringdal, 2018). Uavhengige T-tester ble anvendt når testscore på nominale og uavhengige variabler som for eksempel kjønn og trinn ble undersøkt. Ettersom jeg ikke har utformet hypoteser om for eksempel gutter vil score høyere eller lavere på testen enn jenter, bør det anvendes en tosidig test. En fordel med å bruke T-tester som analyseverktøy er at den gir svar på om det statistisk signifikante forskjeller mellom gjennomsnittet for gruppene (Pallant, 2016). Hvis det er forskjell i gjennomsnitt mellom grupper i utvalget, vil p-verdien vise hvor stor sjans det er for at forskjellen oppsto tilfeldig, og dermed kunne indikere om forskjellene bør generaliseres for hele populasjonen. Før man gjennomfører en t-test bør man bestemme et

signifikansnivå, som p-verdien må være lavere enn for å forkaste nullhypotesen (Ringdal, 2018). Et lavt signifikansnivå reduserer sjansen for å gjøre en type I-feil. Det vanligste signifikansnivået er 0.05 (5%) (Bjørndal & Hofoss, 2004), og anvendes i denne oppgaven. Siden utvalget ikke nødvendigvis er representativt vil ikke p-verdien være gyldig, og resultatene fra undersøkelsen kan derfor uansett ikke generaliseres for populasjonen. Størrelsen på p-verdien gir likevel en indikasjon på om forskjellen i gjennomsnitt er så stor at den bør undersøkes i videre studier eller ikke. Hvor stor forskjellen på gruppene er kan ikke avgjøres av p-verdien, men det kan effektstørrelse gi indikasjoner på. Pallant (2016) anbefaler å måle effektstørrelse i T-tester med Cohens d. Cohens d forteller hvor stor forskjell det er mellom utvalgenes gjennomsnitt, som et standardisert mål. Store forskjeller mellom gruppene gir verdier på Cohens d over 0,8 (Pallant, 2016). Ettersom utvalget er et relativt lite bekvemmelighetsutvalg er små ( $d > 0,2$ ) og moderate ( $d > 0,5$ ) effektstørrelser mindre interessante å undersøke i dette studiet. Variansanalyse (ANOVA) ble anvendt når det var mer enn to grupper som ble sammenlignet, slik Bjørndal og Hofoss (2004) anbefaler.

For å undersøke korrelasjoner ble Pearsons korrelasjonstest anvendt, og slike tester gir verdier av produktmomentkorrelasjonskoeffisient (Pearsons r). Med Pearsons korrelasjonstest kunne jeg undersøke om scorer på i de ulike kategoriene av testen korrelerte. Et eksempel på bruksområdene for dette er å undersøke om elever som scorer høyt på kategoriene og delferdighetene i testen er de samme som fikk en høy score på ferdighetstesten som helhet. Verdien av Pearsons r indikerer hvor sterk korrelasjonen er, og hvilken retning korrelasjonen har (Bjørndal & Hofoss, 2004). En perfekt positiv korrelasjon vil gi  $r = 1$ , mens en perfekt negativ korrelasjon vil gi  $-1$ . Ifølge Pallant (2016) kan  $r > 0,5$  anses tydelig korrelasjon, mens  $r < 0,3$  bør anses som en svak korrelasjon. Korrelasjoner mellom testscore og delferdighetene ble undersøkt for å undersøke om det var de samme elevene som viste generelt høyt ferdighetsnivå gjennom testen, og scoret høyt på de enkelte delferdighetene og testkategoriene.

For å undersøke om det er korrelasjon mellom holdningene og ferdigheter innen kritisk tenkning ble Spearmans rang-korrelasjonstest anvendt. Denne testen egner seg for å undersøke ordinale verdier, som holdningskategoriene er målt i (Pallant, 2016). Testen gir en  $r_s$ -verdi, som er en koeffisient i likhet med Pearsons r. Begge korrelasjonstestene gir i tillegg en p-verdi, som forteller hvor signifikante resultatene er, men de samme forbeholdene om generalisering som ble beskrevet i forrige avsnitt er gjeldene også her.

### 3.4 Ethiske hensyn

Alle studier på mennesker har ett etisk aspekt, som forskere må anerkjenne og legge til rette for at forskningen ikke kan skade respondentene. Forskningsetikken er ment for å beskytte respondentene ved at forskeren skal følge etiske retningslinjer og regler (Ringdal, 2018). Å anvende egen arbeidsplass til forskning gjør at man bør tenke nøye gjennom hvordan prosjektet kan påvirke respondentene. Jeg forsøkte å være nøye med å informere elevene om at det var valgfritt å delta i undersøkelsen, og at jeg hadde full forståelse dersom noen valgte å ikke delta. Ett av de viktigste etiske hensynene innen forskning på mennesker er å unngå å samle informasjon om deltakerne som kan skade dem. Et grep for å gjøre dette er å kun samle inn informasjon som er relevant for studiens hensikt. Jeg har forsøkt å lage variabler som ikke er inngripende og som gjør det vanskelig å gjenkjenne respondentene ut fra datamaterialet. I tillegg har jeg forsøkt å i hovedsak fremstille elevenes svar som statistisk informasjon, slik at det ikke er mulig å gjenkjenne elevene. Enkelte elevsitat er gjengitt i oppgaven, men de innehar ingen informasjon om eleven, og er derfor umulig å knytte til elevens identitet. Hoveddelen av masterprosjektet, altså studie 3 og 4, ble undersøkt anonymt. Utvalg 3 og 4 ble rekruttert gjennom naturfagslæreren deres. Informasjon om undersøkelsen ble gitt muntlig av naturfagslæreren noen dager før og rett før undersøkelsen. I tillegg ble det utsendt en digital skolemelding til foresatte og elever, med informasjon om studien. Siden undersøkelsen er anonym godkjente NSD at det ikke var behov for samtykke fra foresatte, gitt at foresatte ble informert om undersøkelsen på forhånd, og elevene fikk muligheten til å reservere seg mot å delta. Samtykket ble altså samlet inn fra elevene, og hvis de ikke ønsket å delta trengte de ikke å trykke seg videre fra første side i spørreundersøkelsen. De elevene som ønsket å reservere seg fra undersøkelsen kunne arbeide med et alternativt undervisningsopplegg. Testen som utvalg 1 gjennomførte var ikke anonym. Utvalg 1 går i en klasse på 9. trinn som jeg underviser selv i naturfag, og undersøkelsen ble gjennomført i forbindelse med at vi arbeidet med temaet kritisk tenkning. Jeg informerte elevene om prosjektet og kartleggingsverktøyet i forkant av undersøkelsen. Siden alle informantene i utvalg 1 er under 15 er det deres foresatte som må underskrive på samtykkeskjemaene (NSD, u.å.-a), men elevenes samtykke ble innhentet i tillegg. Elevenes samtykke ble etterspurt for å være sikre på at de var fortrolige med at svarene på undersøkelsen kan gjengis i oppgaven. Det ble også innhentet samtykke fra utvalget i studie 2. Behovet for samtykkeskjema kommer av at undersøkelsen ikke ble gjennomført anonymt, i motsetning til resten av studien. Derfor kreves det strengere personverns hensyn (NSD, u.å.-b).

## 4 Resultater

Dette kapittelet vil omhandle resultatene, analysene og endringene som har blitt gjort på kartleggingsverktøyet gjennom prosjektets ulike studier, med hensikt om å besvare studiens to forskningsspørsmål. Kartleggingsverktøyet har gjennomgått revideringer, med utgangspunkt i resultatene fra studie 1, 2 og 3, og mulige endringer vil foreslåes fra studie 4. Studie 1 til 3 gjennomgås først i kapittelet, og har fokus på å videreutvikle og kvalitetssikre ferdighetstesten. Holdningskategoriene vil undersøkes først fra studie 3. Ettersom prosjektet er basert på en stegvis testutvikling, vil endringer og tolkninger drøftes i dette kapittelet. Grunnen til dette er at resultatene på senere studier er avhengige av endringer og tolkninger på foregående studier. Grunnet prosjektets omfang har det ikke vært mulig å løfte frem alle resultater, tolkninger og endringer i dette kapittelet, men kun det som er mest sentrale for testutviklingen.

### 4.1 Studie 1

Studie 1 ga et omfattende datamateriale om elevenes forståelse av kritisk tenkning og den tilhørende versjonen av testen, og utvalgte funn fra studie 1 blir presentert i dette delkapittelet. Hvis det er ønskelig å en fullstendig oversikt over endringene som ble gjort på ferdighetstesten er det mulig å sammenligne versjonen av testen som ble anvendt i studie 1 (se vedlegg 1.2 og 1.3) med testen som ble anvendt i studie 2 (se vedlegg 2.1).

De viktigste funnene ved studie 1 gir innsikt i hvordan de tre ulike delene av testen fungerte, med utgangspunkt i svarene fra de 15 elevene som ønsket å bidra i masteroppgaven. En viktig observasjon fra studie 1 var at omfanget av testen var for stor, ettersom testen var delt opp i to tester, og gjennomført over til sammen to klokke timer. Daværende del 1 (argumentsoppgaver) og del 2 (konklusjonsoppgaver) ble gjennomført som en test, mens en egen test om kildekritikk (del 3) ble gjennomført tre uker senere. Den gjennomsnittlige poengsummen for hele studie 1 var 32 av 49 mulige poeng. Vanskelighetsgraden for hele testen var  $p=0,67$ , som betyr at elevene klarte gjennomsnittlig 67% av oppgavene.

Det ble gjort noen generelle endringer på ferdighetstesten etter studie 1. For det første ble mange av de åpne oppgavene omgjort til lukkede oppgaver som målte det samme. Dette ble gjort for å gjøre kartleggingsverktøyet mindre omfattende, enklere å vurdere og for å øke reliabiliteten. Testen ble også tilpasset UiO nettskjema, slik at videre testing kan skje

anonymt. I tillegg muliggjør nettskjema konvertering til Excel, og enklere omkoding av variabler før datamaterialet eksporteres til SPSS. Det ble også utformet en retteveiledning med kriterier for poenguttelling på de tre åpne oppgavene som ble videreført til studie 3 (se vedlegg 4.2).

#### 4.1.1 Del 1: Argumentsoppgaver

Den største endringen fra studie 1 ble gjort på argumentasjonsoppgavene i del 1. Mye tydet på at argumentasjonsoppgavene ikke diskriminerte tilfredsstillende mellom de elevene som var kompetente til å vurdere argumenter ut fra kriterier og de som ikke var det. Den gjennomsnittlige vanskelighetsgraden for del 1 var lav ( $p=0,70$ ). Dette indikerte at argumentasjonsoppgavene generelt var for lite utfordrende for å skille godt mellom ferdighetsnivået til respondentene. Denne delen besto av 14 AC-oppgaver hvor elevene skulle vurdere om et argument er sterkt eller svakt ut fra kriterier og en åpen kortsvarsoppgave, hvor de skulle forklare hvordan de tenkte. Vanskelighetsgraden til hver enkelt oppgave i del 1 kan ses i Tabell 4-1. Det er interessant at alle oppgavene med sterke argumenter hadde lav vanskelighetsgrad. Oppgave 2 ( $p=0.20$ ) og 11 ( $p=0.33$ ) var de eneste oppgavene hvor flertallet av elevene tok feil, og begge argumentene var svake. Det som var felles for argumentene var at de ikke var relevante for spørsmålet i casen, og inneholdt mye tekst.

Tabell 4-1: Oversikt over vanskelighetsgraden til argumentasjonsoppgavene.

Argument- oppgave	Vanskelighets- grad (p)	Argument (fasit)
1	0,80	Ja, det bør han (svakt)
2	0,20	Ja, for solcellepanel er en klimavennlig energikilde. Det er viktig at vi velger solcellepanel og andre fornybare energikilder for å stoppe global oppvarming. (svakt)
3	0,80	Ja, for de som har solcellepanel på taket har gjennomsnittlig bedre råd. (svakt)
4	0,93	Nei, han vil ikke spare penger på å kjøpe solcellepanel fordi solcellepanel er dyrt, og har lav gjennomsnittlig levetid. Den går altså ofte i stykker før man har tjent inn det solcellepanelene kostet. (sterkt)
5	0,80	Ja, solcellepanel er en god investering. I løpet av fem år vil han ha spart like mye penger på strømutfgifter som han brukte på å kjøpe solcellepanelene. Etter det vil han fortsette å spare penger. (sterkt)
6	0,93	Ja, banken til Ola sier at solcellepanel er en god investering (svakt)
7	0,87	Ja, mange av som har solcellepanel er veldig fornøyde med dem (svakt)
8	0,87	Nei, det synes jeg ikke. (svakt)
9	0,87	Ja, ved å bruke genteknologi til å kurere sykdommer vil det bli færre syke i samfunnet, og man kan bruke tiden på å behandle andre syke pasienter på sykehusene. (sterkt)
10	0,60	Ja, man bør alltid jobbe for å forbedre naturen. Samfunnet vil bli bedre fordi genteknologi kan gjøre at mennesker blir bedre på jobbene sine. (svakt)
11	0,33	Nei, å forandre på naturen er ikke riktig. Hvis man begynner å forandre på den genetiske oppskriften til dyr og mennesker, vil det bare være et spørsmål om tid før det ikke er noen grenser på hvilke endringer vi har gjort på mennesker. (svakt)
12	0,87	Nei, den genetiske oppskriften til mennesker er komplisert, hvis vi endrer på en egenskap vil vi i mange tilfeller også endre på andre egenskaper hos mennesker, som vi ikke vil endre på. De som blir behandlet med genteknologi kan derfor bli syke av behandlingen (sterkt)
13	0,93	Ja, en kjent professor som heter Audun Smart ser ingen utfordringer med å bruke genteknologi for å helbrede syke. (svakt)
14	0,67	Nei, for enten må vi tillate at vi kan forandre på alle egenskaper hos mennesket, eller så må vi forby det. Å kun helbrede syke er ikke mulig. (svakt)
Gj. p-verdi	0,70	

Etter argumentasjonsoppgavene i testen var det en kortsvarsoppgave (oppgave 15) hvor elevene skulle gjøre rede for hvordan de tenkte når de vurderte argumentene. Under halvparten av elevene nevnte at de brukte kriteriene for å vurdere argumentene. Det var kun to elever som nevner alle kriteriene for et sterkt argument i sin besvarelse, og som viste god kriteriebevissthet om hva som kjennetegner sterke argumenter. En av elevene svarte slik:

- «Jeg tenkte over om svarene utrykte en mening og ikke fakta og om det ikke brukte riktig informasjon eller ikke. Jeg så også om det var direkte relevant til spørsmålet eller ikke. Noen var vanskeligere, fordi man skrev relevant informasjon, men den utrykte en mening og brukte ikke riktig informasjon.»

Fire elever nevner enkelte kriterier for sterke argumenter, men nevner ikke at et sterkt argument må være relevant for spørsmålet. Resten av elevene nevnte ikke noen av kriteriene, og flere svarte at de *«tok det som hørtes riktig ut»*, eller at de fokuserte på om argumentet *«var sant eller ikke»*.

Flertallet av argumentasjonsoppgavene hadde altså lav vanskelighetsgrad, men samtidig var det et fåtall av elevene som redegjorde for at de brukte kriteriene for sterke og svake argumenter når de løste oppgavene. Målet var at kun de kompetente elevene, som vurderte argumentene ut fra kjennetegnene for sterke og svake argumenter, skulle få uttelling på oppgavene. Slik oppgavene var i studie 1 kan det virke som det var for enkelt å gjette riktig, eller å vurdere argumentene uten å forholde seg til kjennetegnene. Det var to muligheter for å motvirke dette. Det ene var å øke antallet oppgaver, og det andre var å øke antallet sannsynlige distraktorer. Av hensyn til hvor omfattende kartleggingsverktøyet var, ble antallet distraktorer økt til fire per oppgave. Dette ble gjort ved å dele svaralternativet «svakt argument» inn i tre svaralternativer. Et eksempel på et slik alternativ er: *«Argumentet er svakt, fordi begrunnelsen ikke er relevant til spørsmålet»*. Det ble gjort enkelte andre endringer, som å for eksempel endre argumentet i oppgave 14, som en elev poengterte var logisk problematisk. Argumentet ble endret slik at tolkningsrammen ble mindre.

#### 4.1.2 Del 2: Konklusjonsoppgaver

I del 2 av testen ble eleven presentert for en case hvor de skulle finne den mest sannsynlige forklaringen på hvorfor det har blitt mindre tang og tare i kysten på utsiden av Alaska. De ble presentert for en rekke mulige forklaringer, som de måtte utelukke. Denne delen besto av 9 oppgaver, og elevene fikk gradvis mer informasjon om casen gjennom oppgavesettet. Den første oppgaven i oppgavesettet spurte elevene om global oppvarming kan være årsaken til at det har blitt mindre tang og tare. Kun 9 elever svarer at dette er mulig. Dette kan tyde på at enkelte av elevene har misforstått oppgaven, ettersom det ikke er gitt informasjon som utelukker forklaringen. En elev var allerede helt klar for å konkludere etter å ha blitt presentert for denne forklaringen: *«Det mangler ikke informasjon fordi det er en sammenheng med hva skjedde og konsekvenser.»*. Global oppvarming som forklaring ble utelukket av informasjonen som ble gitt gjennom oppgavesettet, men 6 elever valgte likevel at global oppvarming var den mest sannsynlige forklaringen på casen. Som vist i Tabell 4-2 var det kun 3 elever som valgte den riktige forklaringen.

Tabell 4-2: Fordeling av svar på «den mest sannsynlige forklaringen»

Antall elever som valgte forklaringen	Forklaring
6	Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
3	Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
3	Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere. (Riktig forklaring)
1	Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
1	Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
1	Plastikk fører til at fiskebestanden har gått ned.

Det virket som om mange av elevene ble overveldet av mengden informasjon i denne delen, og at de manglet strategier for å utelukke forklaringer ved å studere informasjonen de ble gitt. En elev svarte for eksempel «Jeg tok det som hørtes rett ut.», og en annen svarte «Jeg gjettet» når de skulle forklare hvordan de hadde tenkt. Noen elever viste at de strategisk vurderte informasjonen om casen, og forsøkte å utelukke mulige forklaringer. En av dem som fant riktig forklaring formulerte strategien sin slik: «Jeg utelukket alternativene ved å se på den informasjonen jeg hadde. Alle de andre alternativene stemte ikke med informasjonen, så da var det bare en igjen og det var svaret.»

Det ble gjort noen endringer etter studie 1. For det første ble det endret på informasjonen elevene ble presentert for, og lagt til generell informasjon om økosystemer, rovdyr og byttedyr. Manglende kunnskap om økosystemer kan ha ført til en feiloppfatning om at havotere kan bli planteetere, og derfor spise tang og tare. Tabell 4-2 viser at tre elever valgte denne forklaringen, i tillegg til at en elev reflekterte rundt dette i en åpen oppgave.

Det også ble lagt til en graf i oppgavesettet, slik at elevene får flere ulike former for informasjon å belage seg på når de skal velge den mest sannsynlige forklaringen. I tillegg ble kun to åpne oppgaver videreført, og disse undersøkte om elevene kunne formulere alternative forklaringer, og gjøre rede for hvilke strategier de brukte for å finne den sannsynlige forklaringen på casen.

#### 4.1.3 Del 3: Oppgaver om kildekritikk

Del 3 av ferdighetstesten ble gjennomført som en separat test tre uker etter del 1 og 2.

Resultatene på del 3, hvor elevene skulle vise forståelse for og vurdere to utdrag av tekster fra internett, viste at det er stor variasjon på elevenes ferdigheter til kritisk lesning. Etter å ha lest



blogginlegget til Sophie Elise, altså kilde 1, var det 5 av elevene som mente at teksten var troverdig. Teksten fra faktisk.no, altså kilde 2, mente 12 av elevene var troverdig. I en oppgave i del 3 ble elevene bedt om å fortelle hvilken kilde de mente var mest troverdig, og begrunne hvorfor. De fleste mente her at kilde 2 var mest troverdig. De to elevene som mente at kilde 1 var mest troverdig, begrunnet dette med at teksten var lett å forstå, interessant eller hadde gode argumenter. Siste spørsmålet på del 3 etterspør kjennetegn på en troverdig kilde. Det varierte hvor mange kjennetegn elevene ramset opp. De fleste av elevene kunne gjengi få og lite konkrete kjennetegn på troverdige nettsider. Sitatet under er et eksempel på en elev som kunne gjøre rede for flere faktorer for hva som kjennetegner en troverdig nettside:

*«En troverdig kilde burde ha godt formulerte setninger uten skrivefeil. Det burde stå i teksten hvor informasjonen som brukes er hentet fra og det burde være hentet fra noen som kan mye eller er ekspert på temaet. Når teksten peker til forskning er den mer troverdig. Ofte er det lurt å se på når kilden ble skrevet. Hvis det er en gammel kilde kan den inneholde informasjon som ikke gjelder lenger.»*

Den største endringen på som ble gjort på del 3 var å erstatte alle kortsvarsoppgavene med lukkede oppgaver, med unntak av oppgaven som spurte elevene om hva som kjennetegner en troverdig kilde. Tekstutdragene ble også ytterligere kortet ned, for å minske gjennomføringstiden. Det ble lagt til en grafisk illustrasjon med statistisk informasjon om kreftrisikoen ved konsumpsjon av kjøtt og røyking, slik at elevenes evne til å forstå illustrasjoner kunne undersøkes.

Da de tre delene i ferdighetstesten ble sammenslått til én test i nettskjema, valgte jeg å ha oppgavene om kildekritikk først, for at elevene kunne starte med å lese tekstutdragene, og ikke miste motivasjonen når disse kom helt til slutt i testen. Oppgavene om kildekritikk ble altså til del 1, argumentsoppgavene ble del 2 og konklusjonsoppgavene ble del 3, og det var denne rekkefølgen som ble brukt i studie 2.

## 4.2 Studie 2

Kartleggingsverktøyet gjennomgikk mange endringer med utgangspunkt tilbakemeldingene som ble gitt i studie 2. Dette delkapittelet tar for seg de største endringene som ble gjort for å kvalitetssikre kartleggingsverktøyet mellom studie 1 og 3. Det er mulig å undersøke de endringene som ikke nevnes i dette delkapittelet ved å studere ulike utkast av testen (se vedlegg 2.1, 2.2 og 3.1).

Både venne- og lærerutvalget i studie 2 ga tilbakemeldinger om skrivefeil og uklarheter, slik at instruksjer, spørsmål og oppgaver ble enklere å forstå. Mengden instruksjer ble redusert for å øke lesbarheten og klarheten i kartleggingsverktøyet. Lærerutvalget identifiserte begreper som kunne være utfordrende for elevene, og det ble lagt til begrepsforklaringer for begreper som «troverdige», «utsagn» og «faktor». Antallet oppgaver ble nedjustert for å gjøre kartleggingsverktøyet mindre omfattende. Lærerutvalget ga tilbakemeldinger om praktiske utfordringer som kunne oppstå, som at enkelte elever kunne bli stående fast på kortsvarsoppgavene. For å unngå dette ble det lagt til en anmodning om at elevene kun skal bruke 2 minutter på kortsvarsoppgaven hvor dette var mest utfordrende (1.0 a).

Det viste seg at flere av oppgavene som ble revidert etter studie 1 var logisk problematiske, fordi distraktorer kunne tolkes som riktige svar. For eksempel var det utfordrende å lage gode lukkede oppgaver om kildekritikk og forståelse av tekstene. Flere av disse oppgaver ble endret mellom venneutvalget og lærerutvalgets gjennomføring (se vedlegg 2.1 og 2.2). Tilsynelatende virket endringene funksjonelle, ettersom lærerutvalget forsto de reviderte oppgavene (f.eks. oppgave 1.1 d).

Den største endringen som ble gjort på ferdighetstesten i studie 2 var å ekskludere hele del 2, altså argumentasjonsoppgavene, fra testen. Som vist til i kapittel 4.1.1 var argumentasjonsoppgavene problematiske i studie 1, og det ble forsøkt å legge til flere distraktorer på hver oppgave før studie 2. Distraktorene var de samme på alle de 14 oppgavene. Eksempeloppgave 5 viser hvordan oppgaven var formulert når venneutvalget gjennomførte kartleggingsverktøyet.

#### 2.1 g) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, mange av dem som har solcellepanel er veldig fornøyde med panelene.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

Eksempeloppgave 5: Oppgave 2.1 g) fra andre utkast av testen (Se vedlegg 2.1)

Endringene fra studie 1 økte vanskelighetsgraden, men utprøvingen på venneutvalget viste at man kunne tenke riktig rundt oppgaven, og likevel velge en distraktorer, fordi noen av argumentene kunne tolkes slik at de passer til flere av svaralternativene enn den tenkte nøkkelen. Det ble gjort endring på distraktorene før lærerutvalget gjennomførte testen, men enkelte oppgaver var fortsatt logisk problematiske. To av lærerne i lærerutvalget misforsto oppgavene i del 2, og fikk ikke med seg at all informasjon i argumentene skulle anses som sann. De ga også tilbakemelding om at mengden informasjon var for stor, og det var derfor lett å overse viktig informasjon. For å gjøre kartleggingsverktøyet mindre omfattende tok jeg avgjørelsen om å fjerne hele del 2, ettersom denne delen fortsatt var problematisk, og endringene ikke hadde fungert. Dette ble gjort på bekostning av den planlagte rollen argumentasjon skulle ha i ferdighetstesten.

I del 3 ble det gitt tilbakemeldinger om at noen av forklaringene ikke ble utelukket fra informasjonen, og endringer som tydeliggjorde dette ble lagt til. To naturfagslærere misforsto instruksjonene i del 3, og vurderte den rette forklaringen til casen som «*ikke mulig*», fordi den virket usannsynlig. Lærerutvalget ga tilbakemelding om at det var for mange instruksjoner i forkant av del 3, og antallet instruksjoner ble redusert, tydeliggjort og forenklet.

#### 4.3 Studie 3

I dette delkapittelet vil resultatene fra studie 3 fremlegges, og oppgavene, holdningsvariablene og kategoriernes reliabilitet diskuteres, og det vil redegjøres for de viktigste endringene som ble iverksatt før studie 4. Versjonen av kartleggingsverktøyet som ble utprøvd i studie 3 er vedlagt (Se vedlegg 3.1). Resultatene fra nettskjema ble analysert i SPSS, etter å ha blitt eksportert og omkodet. Det ble gjennomført item-analyse på oppgavene med utgangspunkt i de anbefalinger som det er redegjort for i teorikapittelet. Tabell 4-3 gir en kort oppsummering av anbefalingene som er relevante for studie 3 og 4. I tillegg til å analysere ferdighetstesten ble den interne konsistensen i holdningsvariablene undersøkt.

Tabell 4-3: Anbefalinger tilknyttet item-analyse

	Vanskelighetsgrad	Diskrimineringssevne	Indre konsistens
<b>Måles med</b>	Facility score (p)	Pearsons r (r)	Cronbach Alfa ( $\alpha$ )
<b>Anbefaling om enkeltoppgaver</b>	Mellom 0.15 og 0.85	$\leq 0.3$	
<b>Anbefalinger om hel test</b>	Gjennomsnittlig p-verdi = 0.5		$\leq 0.65$

Av utvalget på 11 elever var det ingen av elevene som var til stede som trakk seg fra undersøkelsen. Likevel er det begrenset hvor mye troverdighet resultatene fra studie 3 bør gis, ettersom Cohen et al. (2007) anvende mellom 50-100 respondenter hvis man skal undersøke et måleinstruments kvalitet. Det ble derfor ikke fjernet oppgaver fra kartleggingsverktøyet, men oppgavene og kategoriene som ikke nådde kravene for vanskelighetsgrad, diskrimineringssevne eller intern konsistens ble undersøkt nærmere. Endringer på variablene ble gjort når mulig årsak til feilkilde ble identifisert, men kun utvalgte endringer er gjort rede for i delkapittelet. I tillegg ble løsningsforslaget (Se vedlegg 4.2) for testens tre åpne oppgaver utprøvd, og det ble ikke foretatt noen endringer på denne i studie 3.

#### 4.3.1 Ferdighetstesten

Resultatene knyttet til ferdighetstesten som helhet presenteres først. Videre gjøres det rede for resultatene på item-analysen for delferdighetene og utvalgte oppgaver. Den gjennomsnittlige vanskelighetsgraden for hele ferdighetstesten ble målt til 0,66 i studie 3, som vil si at testen er enklere enn optimalt for å differensiere elevenes ferdighetsnivå (McAlpine, 2002). Tabell 4-4 viser at det var intern konsistens ( $\alpha=0,80$ ) i ferdighetstesten som helhet. Som det vil gjøres rede for under var det likevel ikke alle delferdighetene som ga konsistente mål.

Tabell 4-4: Nøkkelinformasjon om ferdighetstesten fra studie 3

	<b>Cronbach Alfa (<math>\alpha</math>)</b>	<b>Vanskelighetsgrad (p)</b>	<b>Gj. Snitt</b>	<b>Max score</b>	<b>Oppgaver (n)</b>	<b>Respondenter (n)</b>
<b>Hel test</b>	0,80	0,66	36,27	55	51	11

Delferdigheten «strategier» (S1) ble målt gjennom to åpne oppgaver. Diskrimineringssevne og vanskelighetsgrad på oppgavene var akseptable, men som Tabell 4-5 viser var den interne konsistensen svært lav ( $\alpha= -0,11$ ). Det ble ikke foretatt endringer, men ble undersøkt nærmere i studie 4.

Tabell 4-5: Oversikt over item-analyse fra delferdigheten S1

<b>Delferdighet</b>	<b>Oppgave (Endret til)</b>	<b>Type oppgave</b>	<b>Vanskelighetsgrad (p)</b>	<b>Diskrimineringssevne (d)</b>	<b>Cronbach Alfa hvis slettet</b>
S1 ( $\alpha= -0,11$ )	1.0 a	Åpen	0,58	0,48	- (kun 2 oppgaver)
	2.5 d (2.4 c)	Åpen	0,37	0,67	- (kun 2 oppgaver)

Delferdigheten «Vurdere troverdighet» (S2) hadde en lav gjennomsnittlig vanskelighetsgrad ( $p=0,75$ ), og den interne konsistensen var under akseptabelt nivå ( $\alpha=0,60$ ). Som det vises i

Tabell 4-6, fikk alle elevene til 1.1 i) og 1.1 j), og disse oppgavene kan være for enkle for å gi innsikt i elevers kompetanse. Derfor ble oppgave 1.1 i) endret, for å øke vanskelighetsgraden. Oppgave 1.2 h) bidro negativt til den interne konsistensens og hadde negativ diskrimineringssevne. Oppgaven ble endret fra «Påstandene i kilde 2 stemmer med det de fleste tror om kjøtt og røyking fra før, og dette styrker troverdigheten.» til «Påstandene i kilde 2 er mindre ekstreme enn i kilde 1, og dette øker troverdigheten til kilde 2», slik at ikke respondentens inntrykk av «folk flest» påvirker. Alle oppgavene i denne delferdigheten, unntatt en, har kun en distraktor og en nøkkel, og var av typen AC-oppgaver. Dette kan forklare den lave vanskelighetsgraden i kategorien. Dette vil undersøkes nærmere i studie 4.

Tabell 4-6: Oversikt over item-analyse fra delferdigheten S2

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskrimineringssevne (d)	Cronbach Alfa hvis slettet
S2 ( $\alpha=0,60$ )	1.1 a	Lukket (2m)	0,73	0,61	0,53
	1.1 b	Lukket (4m)	0,45	0,82	0,44
	1.1 g (1.1 e)	Lukket (2m)	0,45	0,46	0,59
	1.1 h (1.1 f)	Lukket (2m)	0,73	0,51	0,57
	1.1 i (1.1 g)	Lukket (2m)	<b>1</b>	-	-
	1.1 j (1.1 h)	Lukket (2m)	<b>1</b>	-	-
	1.1 l (1.1 i)	Lukket (2m)	0,64	0,57	0,55
	1.2 a	Lukket (2m)	<b>0,91</b>	0,32	0,60
	1.2 e	Lukket (2m)	0,82	0,71	0,51
	1.2 g	Lukket (2m)	0,82	0,35	0,60
	1.2 h	Lukket (2m)	0,55	<b>-0,18</b>	<b>0,74</b>
	1.3 a	Lukket (2m)	<b>0,91</b>	0,63	0,54

Delferdigheten N1 måler elevenes evne til å avkode og tydeliggjøre informasjon fra tekster.

Tabell 4-7 viser at verdien av Cronbach Alfa ( $\alpha=0.45$ ) er lavere enn akseptabelt nivå.

Oppgave 1.1 d) og 1.2 b), bidrar negativt til den interne konsistensen i delferdigheten, og har lav diskrimineringssevne og vanskelighetsgrad. Nøkkelen i oppgave 1.1 d) ble endret fordi den kunne være enkel å identifisere, for å øke vanskelighetsgraden. Det ble ikke funnet en mulig feilkilde ved 1.2 b), og tyder på at de fleste elevene forsto budskapet i kilde 2. Oppgave 1.2 c) diskriminerer godt, men har lav vanskelighetsgrad. Det ble gjort endringer på distraktorene for å øke vanskelighetsgraden.

Tabell 4-7: Oversikt over item-analyse fra delferdigheten N1

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskriminerings-evne (d)	Cronbach Alfa hvis slettet
N1 ( $\alpha=0,45$ )	1.1 c	Lukket (4m)	0,73	0,81	0,10
	1.1 d	Lukket (4m)	<b>0,91</b>	<b>0,09</b>	<b>0,62</b>
	1.2 b	Lukket (4m)	<b>0,91</b>	<b>0,09</b>	<b>0,62</b>
	1.2 c	Lukket (6m)	<b>0,91</b>	0,75	0,19
	1.2 d	Lukket (4m)	0,82	0,87	-0,51

Det er en høy interne konsistens i delferdigheten N2 ( $\alpha=0.80$ ), som måler elevenes evne til å avkode informasjon fra tabeller og illustrasjoner. Tabell 4-8 viser at det var lav vanskelighetsgrad i oppgavene, med unntak av 1.4 d). Alle oppgavene har høy diskrimineringssevne, utenom 1.4 b) som alle fikk til. Alle oppgavene beholdes uendret.

Tabell 4-8: Oversikt over item-analyse fra delferdigheten N2

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskriminerings-evne (d)	Cronbach Alfa hvis slettet
N2 ( $\alpha=0,80$ )	1.4 a	Lukket (4m)	<b>0,91</b>	0,77	0,74
	1.4 b	Lukket (5m)	<b>1</b>	-	-
	1.4 c	Lukket (5m)	0,82	0,83	0,71
	1.4 d	Lukket (5m)	0,36	0,69	<b>0,81</b>
	2.5 a	Lukket (4m)	0,82	0,67	0,78
	2.5 b	Lukket (4m)	0,73	0,81	0,73

Delferdigheten K1 måler om elevene får til å vurdere om ulike konklusjoner er mulige basert på bevis. Det er lite intern konsistens i kategorien ( $\alpha=0.23$ ). Som det vises i Tabell 4-9 hadde oppgavene tilknyttet delferdigheten generelt høy vanskelighetsgrad. Det er problematisk at fire av oppgavene har lav diskrimineringssevne, og bidrar negativt til den interne konsistensen. Den faktiske løsningen på casen om hvorfor tang og tare forsvinner fra havbunnen utenfor Alaska, er det få som velger at «er mulig», selv for de elevene som svarer korrekt på de andre oppgavene. Det kan virke som at elevene synes forklaringen er usannsynlig, selv om den ikke kan utelukkes fra informasjonen de har blitt gitt. Instruksjonen «En forklaring er mulig, hvis den ikke utelukkes fra informasjonen» ble lagt oppgavens stamme, slik at ikke de kompetente elevene utelater denne forklaringen.

Tabell 4-9: Oversikt over item-analyse fra delferdigheten K1

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskriminerings-evne (d)	Cronbach Alfa hvis slettet
K1 ( $\alpha=0,23$ )	2.1 a	Lukket (2m)	0,55	0,56	0,08
	2.2 a	Lukket (2m)	0,55	0,56	0,08
	2.2 b	Lukket (2m)	0,82	<b>0,14</b>	<b>0,27</b>
	2.2 c	Lukket (2m)	0,73	0,40	0,17
	2.2 d	Lukket (2m)	0,18	<b>-0,39</b>	<b>0,41</b>
	2.2 e	Lukket (2m)	0,36	0,49	0,16
	2.2 f	Lukket (2m)	0,36	<b>-0,42</b>	<b>0,48</b>
	2.3 b	Lukket (2m)	0,73	0,50	0,11
	2.3 c	Lukket (2m)	0,55	0,75	-0,08
	2.3 d	Lukket (2m)	0,64	0,42	0,16
	2.3 e	Lukket (2m)	0,27	<b>-0,40</b>	<b>0,45</b>
	2.3 f	Lukket (2m)	0,27	0,48	0,13
	2.3 g	Lukket (2m)	0,55	0,75	-0,08

Oppgavene i delferdigheten K2 måler elevenes evne til å vurdere bevisgrunnlag. Elevene fikk spørsmål om når det er tilstrekkelig informasjon for å ta avgjørelse i lukkede oppgaver, og på en åpen oppgave ba elevene om å formulere egne alternative forklaringer. Som Tabell 4-10 viser er det variert vanskelighetsgrad i oppgavene, og diskrimineringssevnen til oppgavene er akseptabel, men til tross for dette er det ikke intern konsistens i delferdigheten ( $\alpha=0,07$ ). Det er få oppgaver tilknyttet kategorien, og sammensetningen av en åpen og tre AC-oppgaver kan være årsaken til dette. Det ble foretatt endringer for å øke klarheten i 2.1 b), 2.3 a) og 2.4 a).

Tabell 4-10: Oversikt over item-analyse fra delferdigheten K2

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskriminerings-evne (d)	Cronbach Alfa hvis slettet
K2 ( $\alpha=0,07$ )	2.1 b	Lukket (2m)	0,73	0,36	<b>0,13</b>
	2.1 c	Åpen	0,59	0,79	-0,53
	2.3 a	Lukket (2m)	0,55	0,49	0,03
	2.4 a	Lukket (2m)	0,55	0,32	<b>0,26</b>

Delferdigheten K3 måler elevenes evne til å trekke nøye gjennomtenkte konklusjoner, med grunnlag i bevis. Tabell 4-11 viser at det er lite intern konsistens i delferdigheten ( $\alpha= -0,44$ ). Mulige årsak kan være kategoriens sammensetning av oppgaver fra både fra del 1 og del 2 av testen, med ulikt antall distraktorer. Oppgave 2.4 b bidrar negativt til den interne konsistensen, men om den ble omplassert kunne den påvirke den interne konsistensen i K1

positivt. Hvis utprøvingen i studie 4 viser like lav korrelasjon mellom oppgavene, bør det foretas endringer på testkategoriene

Tabell 4-11: Oversikt over item-analyse fra delferdigheten K3

Delferdighet	Oppgave (Endret til)	Type oppgave	Vanskelighetsgrad (p)	Diskriminerings-evne (d)	Cronbach Alfa hvis slettet
K3 ( $\alpha = -0,44$ )	1.3 b	Lukket (2m)	<b>0,91</b>	0,43	-0,61
	1.3 c	Lukket (2m)	<b>0,91</b>	0,43	-0,81
	2.4 b	Lukket (6m)	0,36	0,39	<b>0,23</b>
	2.5 c	Åpen	0,73	0,53	- 0,30

Det ble i tillegg utformet en delferdighet (K4) som er tilknyttet testkategorien K, hvor elevene skal vurdere argumenter, for å øke fokuset på argumentasjon i testen. Argumentsoppgavene er mindre omfattende og tekststunge enn de som ble fjernet etter studie 2. Oppgave 1.1 j) ble flyttet til denne delferdigheten og oppgave 1.3 f-g) ble lagt til testen.

#### 4.3.2 Holdningskategoriene

Det er vedlagt en oversikt over den interne konsistensen i holdningskategoriene for studie 3 (Se vedlegg 3.2). Hver holdningskategori besto av fire spørsmål, med unntak av «holdninger til viktigheten av kritisk tenkning i skolen», som ble målt gjennom tre spørsmål.

Holdningskategorien «holdninger til naturfag» har en høy intern konsistens ( $\alpha=0.81$ ), og derfor ble det ikke gjort endringer på denne. Det ble ikke gjort endringer på kategoriene «holdninger til viktigheten av kritisk tenkning i skolen» ( $\alpha=0.59$ ) og «sannhetssøkende» ( $\alpha=0.51$ ), fordi den interne konsistensen var nesten på akseptabelt nivå, og utvalgets størrelse gir store feilmarginer. Det ble gjort endringer på de andre holdningskategoriene ettersom det var lite interne konsistens i kategoriene. De fleste av revideringene av spørsmålene hadde som formål å fjerne flerdimensjonlitet i spørsmålene. Det ble endret på variablene som var inspirert av Facione (2015; 2000), ettersom disse kunne oppfattes som flerdimensjonale. Løsningen ble å forenkle disse, og øke klarheten. Det ble foretatt endringer på spørsmål som bidro negativt til den indre konsistensen som inneholdt ord som «*alltid*» og «*ikke*», fordi det anbefales å unngå disse ordene.



## 4.4 Studie 4

I dette kapittelet presenteres resultatene fra studie 4. Utvalget som ble undersøkt i studie 4 var betraktelig større enn i studie 1, 2 og 3, og derfor legges det mer vekt på resultatene fra denne delen av masterprosjektet. Resultatene knyttet til studiens to forskningsspørsmål vil behandles stegvis. Først gjøres det rede for resultatene knyttet til testens kvalitet (FS1) med utgangspunkt i item-analyse. I tillegg vil resultatene på testen undersøkes, for å lage en oversikt over vanskelighetsgraden til de ulike delferdighetene på testen (FS2.a.) og om det finnes korrelasjoner mellom andre variabler og score på ferdighetstesten (FS2.b.). Det er vedlagt en oversikt over hvilke oppgaver som er tilknyttet de ulike kategoriene og delferdighetene (se vedlegg 4.1). Versjonen av kartleggingsverktøyet<sup>6</sup> som ble anvendt i studie 4 er vedlagt (se vedlegg 4.4).

### 4.4.1 Utvalg, frafall og dataregngjøring

De fire klassene i utvalg 4 har totalt 113 elever. Det var frafall fra utvalget, ettersom noen elever ikke ønsket delta i undersøkelsen. I tillegg ønsket naturfagslærerne å frita enkelte elever, med begrensede norskerferdigheter. Det var høyt fravær når undersøkelsen ble gjennomført, som bør ses i sammenheng med at elever skulle holde seg hjemme ved tegn til sykdom, grunnet Covid-19-situasjonen. Lærerne rapporterte at til sammen 101 elever var til stede når undersøkelsen ble gjennomført, og 88 elever deltok i studien, mens 13 valgte å ikke delta, eller ble fritatt av naturfagslærer.

Det ble lastet opp 76 besvarelser i nettskjema. Antallet leverte besvarelser stemmer dermed ikke overens med antallet respondenter som lærerne rapporterte om. Dette kan indikere at elever bevisst eller ubevisst ikke leverte inn besvarelsene sine via nettskjema. På dagen hvor to av klassene skulle gjennomføre undersøkelsen, ble tilkoblingen til internett brutt for elevene over lengre perioder, og var ustabil gjennom skoledagen. Dette kan være hovedårsaken for de manglende besvarelsene. Spesielt avviker antallet respondenter fra 10. trinn fra rapportert antall, og en klasse på 10. trinn gjennomførte kartleggingsverktøyet samtidig som internett var ustabil. Det er sannsynligvis de elevene som brukte lengst tid på

---

<sup>6</sup> Det er mulig å prøve kartleggingsverktøyet i nettskjema, for å få en bedre innsikt i hvordan undersøkelsen opplevdes for elevene. Resultatene fra testen vil lagres anonymt, men det lagres informasjon om dato for gjennomføring (<https://nettskjema.no/a/kritisk-tenkning>).

undersøkelsen som ikke fikk levert besvarelsene sine i denne klassen, og dette kan ha påvirket resultatene på kartleggingsverktøyet.

Det ble gjennomført en datarensing av de 76 besvarelsene som ble sendt inn. Antallet deltagende respondenter etter datarensingen var 72, ettersom 4 respondenter ble ekskludert fra utvalget, etter å ha gjennomført en skjønsmessig vurdering av datamaterialet. Tre av de fire elevene som ble ekskludert gjennomførte hele kartleggingsverktøyet på svært kort tid (3m 42s, 9m 52s, 7m 37s), og med tanke på kartleggingsverktøyets omfang ble det betraktet som urealistisk at elevene kan ha gjort sitt beste og svart ærlig i sine besvarelser. På besvarelsene som ble fjernet var ikke de åpne oppgavene besvart, eller de svarte useriøst (som «*hæ?*», «*elskling*» og «*jeg liker ost*»). Elevene scoret spesielt lavt på lukkede oppgavene med flere distraktorer, og dette indikerer at de kan ha gjettet uten å lese oppgavene. To av elevene svarte «*helt enig*» på alle holdningsvariablene. En av respondentene hvor besvarelsen ble fjernet fra datagrunnlaget brukte mer tid enn de tre nevnte, men ble fjernet grunnet at respondenten svarte «*helt enig*» på alle holdningsvariablene, og de tre åpne oppgavene ble besvart med useriøst.

#### 4.4.2 Resultater knyttet til forskningsspørsmål 1: Kvalitetssikring av kartleggingsverktøyet

Alle resultater fra dette masterprosjektet er avhengig av kvaliteten på måleinstrumentet. Gjennom foregående kapitler har operasjonaliseringen og kvalitetssikringen av kartleggingsverktøyet blitt gjort rede for. Under redegjøres det for de resultater fra studie 4 som er relevant for testverktøyets reliabilitet og validitet. Testen av ferdigheter vil vies størst fokus i dette delkapittelet, men også resultater knyttet til reliabiliteten på holdningsvariablene fremlegges. Jeg har valgt å fokusere på kriterier for intern konsistens, diskrimineringssevne og vanskelighetsgrad som utgangspunkt for item-analyse i testen, mens kun intern konsistens tallfestes for holdningsvariablene. Resultater knyttet til hver testkategori og delferdighet vil gjennomgås først, og mulige endringer på disse diskuteres. Denne diskusjonen skjer i dette kapittelet slik at neste kapittel kan fokusere på å drøfte reliabiliteten og validiteten til testen som helhet, og hvor god testen er som et mål på de ferdighetene som er nødvendige for å tenke kritisk i naturfag.

##### *Strategier for kritisk tenkning (S)*

Testkategorien «Strategier for kritisk tenkning» (S) tar utgangspunkt i ferdighetene «*Forklare*» og «*Selvregulering*» og delferdigheten «*vurdering av troverdighet*» fra

delphistudiet til Facione (1990a). Kategorien har som formål å gi informasjon om elevene har og kan bruke strategier for å kunne ta en avgjørelse og utøve kildekritikk, og om elevene kan gjøre rede for strategiene de anvender for å fatte avgjørelser i testen. Kategorien består av to delferdigheter, med til sammen 15 oppgaver. Kategorien har sammenlagt en akseptabel intern konsistens ( $\alpha=0,78$ ), som vises i Tabell 4-12.

Tabell 4-12: Oversikt over testkategorien «S»

<b>Kategori: Strategier for kritisk tenkning</b>				
<b>Standardisert Cronbach's Alfa</b>	<b>Vanskelighetsgrad (p)</b>	<b>Gj. snitt</b>	<b>Standardfeil</b>	<b>Antall oppgaver (n)</b>
0,78	0,76	14.46	0,37 poeng	15

Tabell 4-13 viser resultatene fra item-analyse på oppgavene i kategorien. Slike tabeller gir oversikt over hvordan oppgavene diskriminerer i forhold til både kategorien og delferdigheten de er tilknyttet. De resultatene fra item-analysen som er dårligere enn anbefalt er uthevet med i tabellen. Den gjennomsnittlige vanskelighetsgraden var lav ( $p=0,76$ ), som viser at de fleste elevene fikk høy score på oppgavene i kategorien.

Tabell 4-13: Resultat på item-analyse i forhold til testkategori og delferdighet

Delferdighet	Oppgavenr.	Vanskelighetsgrad (p)	Testkategori		For delferdighet	
			Diskriminerings-evne (d)	Cronbach's Alpha if Item Deleted	Diskriminerings-evne (d)	Cronbach's Alpha if Item Deleted
S1 ( $\alpha=0,46$ )	1.0 a)	0,60	0,45	0,70	0,30	.
	2.4 c)	0,57	0,37	0,74	0,30	.
S2 ( $\alpha=0,75$ )	1.1 a)	0,82	0,37	0,71	0,49	0,70
	1.1 b)	0,65	0,41	0,71	0,46	0,70
	1.1 f)	0,78	0,57	0,69	0,63	0,68
	1.1 e)	0,67	0,06	0,74	<b>-0,05</b>	<b>0,78</b>
	1.1 g)	<b>0,92</b>	0,54	0,70	0,56	0,70
	1.1 h)	0,81	0,47	0,70	0,60	0,68
	1.1 i)	0,78	0,49	0,70	0,49	0,70
	1.2 a)	<b>0,94</b>	0,26	0,72	<b>0,22</b>	0,73
	1.2 e)	<b>0,93</b>	0,39	0,71	0,31	0,72
	1.2 f)	<b>0,94</b>	0,16	0,73	<b>0,11</b>	0,74
	1.2 g)	<b>0,88</b>	0,25	0,72	<b>0,17</b>	0,74
	1.2 h)	0,85	0,36	0,71	0,35	0,72
	1.3 a)	<b>0,97</b>	0,44	0,72	0,47	0,72

Delferdigheten «strategier» (S1) viser en høyere intern konsistens ( $\alpha=0,46$ ) enn i studie 3, men er fortsatt ikke innenfor en akseptabel verdi ( $\alpha>0,65$ ). Det tyder på at testen ikke gir et reliabelt svar om elevenes kompetanse i delferdigheten. Den lave verdien kan forklares ved at det kun er to kortsvarsoppgaver som måler delferdigheten, og antallet testledd påvirker Cronbach alfa, i tillegg til at man kan forvente mindre intern konsistens ved bruk av åpne oppgaver. De to oppgavene tilhører ulike deler av testen, og handler om forskjellige temaer, og dette påvirker sannsynligvis den interne konsistensen ytterligere negativt. Oppgave 1.0 a) tilhører del 1 av testen, og undersøker elevenes strategier til å vurdere troverdigheten til informasjon hentet fra nettsider. Oppgave 2.4 c) er tilknyttet del 2 i testen, hvor elevene skal finne ut hvorfor det blir mindre tang og tare på havbunnen utenfor Alaska. Oppgaven spør hvordan elevene tenkte når de vurderte de mulige forklaringene på mysteriet. Oppgavene i S1 diskriminerer likevel på et akseptabelt nivå. Oppgavene bidrar dessuten positivt til både testens og testkategoriens interne konsistens, som styrker forklaringen om at den største

feilkilden er knyttet til antallet oppgaver i delferdigheten. Delferdigheten bør beholdes slik den er til tross for lav intern konsistens, ettersom ekskludering av oppgavene ville fjernet viktige aspekter ved begrepet kritisk tenkning.

Delferdigheten «Å vurdere troverdighet» (S2) har som hensikt å måle om elevene kan vurdere en kildes troverdighet gjennom å undersøke ulike aspekter ved en tekst. Delferdigheten er sammensatt av 13 lukkede oppgaver, hvorav 12 er AC-oppgaver. Det er intern konsistens blant oppgavene i delferdigheten ( $\alpha=0,75$ ), og de fleste oppgavene diskriminerer på et akseptabelt nivå ( $d>0,3$ ). Oppgavene utgjør to tredjedeler av poengutdelingen i testkategorien, men flertallet av oppgavene har lav vanskelighetsgrad, altså måler de samme ferdighetsnivå. Det er mulig å redusere antallet oppgaver tilknyttet delferdigheten, med hensyn til den interne konsistensen, og fordi det er mange oppgaver som måler på samme vanskelighetsgrad.

Avveiningen om hvilke oppgaver som bør fjernes bør vurderes med hensyn til innholdsvaliditet. Oppgave 1.1 a) og 1.2 a) har for eksempel lav vanskelighetsgrad, men er ikke aktuelle å ekskludere fra testen, fordi oppgavene kan gi relevant informasjon om hvordan elevene tenkte når de vurderte kilde 1 og 2 i del 1. Oppgave 1.1 e) kan derimot ekskluderes. Den påvirker den interne konsistensen negativt, og diskrimineringsvnen ( $d= -0,05$ ) viser at det var flere av de kompetente elevene som svarte feil på oppgaven. I oppgaven skal elevene svare om de er uenige eller enige i påstanden «*Kilde 1 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.*», og nøkkelen er «*uenig*». Det er ingen tilsynelatende interessekonflikt for forfatteren i kilde 1, men det kan være at elevene oppfatter henvisningene fra Sophie Elise til Netflix-dokumentaren «*What the health*» som å «*selge et produkt*». Samme spørsmål blir stilt om kilde 2 i oppgave 1.2 e), men denne oppgaven diskriminerer bedre. Å kunne identifisere mulige interessekonflikter er et viktig element i kildekritikk, og spesielt relevant for nettsider. Siden interessekonflikter ikke er relevante for verken kilde 1 eller 2 kan oppgave 1.1 e) og 1.2 e) ekskluderes fra testen. Mange av oppgavene som måler delferdigheten har lav vanskelighetsgrad, og det er mulig å ekskludere oppgave 1.2 f) ( $d=0,11$ ,  $p=0,92$ ) og 1.2 g) ( $d=0,16$ ,  $p=0,83$ ) fra testen, med tanke på at oppgavene heller ikke har god evne til å diskriminere. Hvis de to endringene implementeres, vil antallet oppgaver i delferdigheten redusert med fire oppgaver.

#### *Forståelse for naturfaglig informasjon (N)*

Kategorien som måler «forståelse for naturfaglig informasjon» (N) undersøker elevenes evne til å forstå ulike former for naturfaglig informasjon, og identifisere hvilken rolle informasjonen har i teksten. Kategorien har tatt utgangspunkt i ferdighetene «*tolke*» og

«*analysere*» fra Facione (1990a). Kategorien består av tolv lukkede oppgaver med tre distraktorer og en nøkkel. Ifølge Tabell 4-14 er den interne konsistensen ( $\alpha=0,62$ ) i kategorien under akseptabelt nivå ( $\alpha>0,65$ ). Forklaringen er sannsynligvis at dette er kategorien med færrest tilknyttede oppgaver i testen, og ifølge Tabell 4-15 er det kun fem oppgaver som diskriminerer på et akseptabelt nivå. Verken delferdigheten «*avkode og tydeliggjøre informasjon fra tekster*» (N1) ( $\alpha=0,52$ ) eller «*Avkode informasjon fra tabeller og illustrasjoner*» (N2) ( $\alpha=0,47$ ) har en akseptabel verdi for Cronbach alfa ( $\alpha>0,65$ ), til tross for at den interne konsistensen i N2 var over akseptabelt nivå i studie 3 ( $\alpha=0,80$ ). Score i delferdighetene korrelerer tydelig med score i ferdighetstesten ( $r_{N1}=0,63$ ,  $r_{N2}=0,67$ ), til tross for at dette er den kategorien i testen som gir minst poenguttelling. Det betyr at elevene som får høye scorere i ferdighetstesten, i stor grad også fikk høy score i kategorien.

Tabell 4-14: Oversikt over testkategorien "N"

Kategori: Forståelse for naturfaglig informasjon (N)				
Standardisert Cronbach's Alfa	Vanskelighetsgrad (p)	Gj. snitt	Standardfeil	Antall oppgaver (n)
0,62	0,74	8,93	0,26 poeng	15

Tabell 4-15: Resultat på item-analyse i forhold til testkategori og delferdighet

Del-ferdighet	Oppgavenr.	Vanskelighetsgrad (p)	Hel kategori		For delferdighet	
			Diskriminerings-evne (d)	Cronbach's Alpha if Item Deleted	Diskriminerings-evne (d)	Cronbach's Alpha if Item Deleted
N1 ( $\alpha=0,52$ )	1-1-c	0,78	0,22	0,61	<b>0,28</b>	0,46
	1-1-d	0,63	0,27	0,60	0,31	0,45
	1-2-b	0,74	0,29	0,59	0,31	0,44
	1-2-c	0,78	0,25	0,60	0,31	0,44
	1-2-d	0,75	0,47	0,56	<b>0,22</b>	0,50
N2 ( $\alpha=0,47$ )	1-4-a	<b>0,90</b>	0,32	0,59	0,41	0,37
	1-4-b	<b>0,92</b>	0,24	0,60	<b>0,29</b>	0,42
	1-4-c	<b>0,94</b>	0,26	0,60	0,30	0,42
	1-4-d	0,47	0,16	0,62	<b>0,14</b>	<b>0,49</b>
	1-4-e	0,60	0,36	0,58	<b>0,18</b>	0,46
	2-5-a	0,82	0,18	0,61	<b>0,14</b>	0,47
	2-5-b	0,61	0,27	0,60	<b>0,28</b>	0,40

Delferdigheten som innebærer ferdighetene til å «avkode og tydeliggjøre informasjon fra tekster» (N1) har lav intern konsistens. Tre av fem oppgaver diskriminerer over akseptabelt nivå ( $d > 0,3$ ), mens to oppgaver har en lavere diskrimineringssevne enn akseptabelt. Ingen av oppgavene påvirker den interne konsistensen negativt, og det kan tyde på at hovedårsaken til at verdien på Cronbach's alfa er under akseptabelt nivå er at det er få oppgaver knyttet til delferdigheten.

Den interne konsistensen for delferdigheten «Avkode informasjon fra tabeller og illustrasjoner» (N2) ( $\alpha = 0,47$ ) er i likhet med N1 under akseptabelt nivå. Som vist i Tabell 4-15 er det kun to oppgaver som diskriminerer på et akseptabelt nivå ( $d > 0,3$ ). Oppgave 1.4 d) er den eneste oppgaven som påvirker den interne konsistensen negativt for delferdigheten, i tillegg til at den diskriminerer svakt i forhold til både kategorien og delferdighet. Oppgaven er tilknyttet en stimulus, med en figur med statistikk over kreftforekomst ved røyking og kjøtt. Forklaringen på hvorfor oppgaven bidrar negativt kan være at den ligner oppgave 1.4 c), som spør om «*Hvor stor andel av all lungekreft skyldes røyking, ifølge illustrasjonen?*», men oppgave 1.4 d) spør i stedet for om hvor stor andel «*av alle krefttilfeller*» skyldes at folk spiser kjøtt. Flertallet av elevene deles mellom oppgavens nøkkel, og en distraktor, som ville vært riktig svar hvis spørsmålet var «*hvor stor andel av all tarmkreft skyldes at folk spiser kjøtt*». Ettersom det kan virke som at selv kompetente elever misforstår oppgaven bør oppgave 1.4 d) bør ekskluderes fra testen.

Med hensikt om å revidere og forbedre testen, er det ingen umiddelbar løsning for hvordan man kan øke den interne konsistensen i testkategorien N. To forslag kan likevel skisseres; enten kan kategorien ekskluderes fra testen, men dette skjer på bekostning av innholdsvaliditeten til testen, med hensyn til at kategorien er et sammensatt mål på to ferdigheter fra Facione (1990). Testen skal måle elevens ferdigheter til å tenke kritisk i naturfag, og det påvirker testens umiddelbare validitet å fjerne ferdigheter knyttet til det å tolke og forstå naturfaglig informasjon. God forståelse for naturfaglig informasjon er sentralt for å kunne tenke kritisk i naturfag. Det andre alternativet er å øke antallet oppgaver i kategorien, og endre de oppgavene som diskriminerer svakt. Dette alternativet er bedre om man ønsker å bevare innholdsvaliditeten i testen.

#### *Trekke velbegrunnede konklusjoner (K)*

Elevenes ferdighet til å «trekke velbegrunnede konklusjoner» (K) er en kategori som består av fire delferdigheter, som ble utviklet med utgangspunkt i ferdighetene «*Å trekke konklusjoner*»

og «vurdere» fra Facione (1990a). Tabell 4-16 viser at kategoriens interne konsistens ( $\alpha=0,69$ ) er over akseptabelt nivå ( $\alpha>0,65$ ). Kategorien er testens mest omfattende og består av 26 oppgaver, og nesten halvparten av poenguttellingen i ferdighetstesten er tilknyttet kategorien (27 av 58 poeng). Dette bidrar til at det er en tydelig korrelasjon mellom score i ferdighetstesten og kategori K ( $r=0,91$ ). Blant de tre kategoriene i testen, har kategori «K» høyest vanskelighetsgrad. Til tross for akseptabel intern konsistens på kategorinivå, kan man se i Tabell 4-17, som viser resultatene fra item-analyse, at det er utfordringer med å få reliable målinger på delferdighetsnivå, i tillegg til at flere av oppgavene ikke diskriminerer på akseptabelt nivå ( $d>0,3$ ).

Tabell 4-16: Oversikt over testkategorien "K"

<b>Kategori: Forståelse for naturfaglig informasjon (N)</b>				
Standardisert Cronbach's Alfa	Vanskelighetsgrad (p)	Gj. snitt	Standardfeil	Antall oppgaver (n)
0,69	0,61	16,50	0,48	26



Tabell 4-17: Resultat på item-analyse i forhold til testkategori og delferdighet

Delferdighet	Oppgavenr.	Vanskelighetsgrad (p)	Hel kategori		For delferdighet	
			Diskriminerings- evne (d)	Cronbach's Alpha if Item Deleted	Diskriminerings- evne (d)	Cronbach's Alpha if Item Deleted
K1 ( $\alpha=0,24$ )	2.1 a)	0,83	0,23	0,67	<b>0,12</b>	0,21
	2.2.a)	0,79	0,32	0,67	0,33	0,13
	2.2.b)	0,63	0,25	0,67	<b>0,14</b>	0,20
	2.2 c)	0,54	0,26	0,67	<b>0,18</b>	0,18
	2.2 d)	0,38	-0,19	<b>0,71</b>	<b>-0,20</b>	<b>0,34</b>
	2.2 e)	0,38	0,20	0,68	<b>0,11</b>	0,21
	2.2 f)	0,57	-0,14	<b>0,71</b>	<b>-0,19</b>	<b>0,34</b>
	2.3 b)	0,76	0,28	0,67	<b>0,17</b>	0,19
	2.3 c)	0,60	0,26	0,67	<b>0,15</b>	0,20
	2.3 d)	0,58	0,40	0,66	<b>0,17</b>	0,19
	2.3 e)	0,47	-0,07	<b>0,70</b>	<b>-0,10</b>	<b>0,31</b>
	2.3 f)	0,33	0,09	0,69	<b>0,10</b>	0,22
2.3 g)	0,64	0,29	0,67	<b>0,14</b>	0,20	
K2 ( $\alpha=0,27$ )	2.1 b)	<b>0,90</b>	0,18	0,68	<b>0,29</b>	0,07
	2.1 c)	0,54	0,43	0,65	<b>0,26</b>	-0,11
	2.3 a)	0,60	0,32	0,66	<b>0,09</b>	0,22
	2.4 a)	0,65	0,18	0,68	<b>-0,08</b>	<b>0,40</b>
K3 ( $\alpha=0,66$ )	1.3 b)	<b>0,90</b>	0,26	0,67	<b>0,27</b>	<b>0,69</b>
	1.3 c)	<b>0,89</b>	0,24	0,67	<b>0,25</b>	<b>0,69</b>
	2.4 b)	0,47	0,62	0,64	0,67	0,39
	2.5 c)	0,49	0,56	0,64	0,63	0,43
K4 ( $\alpha=0,45$ )	1.1 j)	0,72	0,23	0,68	<b>0,23</b>	0,40
	1.3 d)	0,68	0,31	0,67	0,31	0,34
	1.3 e)	0,40	0,13	0,68	<b>0,29</b>	0,35
	1.3 f)	0,40	0,26	0,67	0,36	0,30
	1.3 g)	0,82	0,23	0,67	<b>-0,01</b>	<b>0,53</b>

Delferdigheten «undersøke og vurdere bevis» (K1) har lav verdi for interne konsistens ( $\alpha=0,24$ ), og kun oppgave 2.2 a) diskriminerer på et akseptabelt nivå, mens resten av

oppgavene ikke har en akseptabel diskrimineringssevne ( $d < 0,3$ ). Det er tre oppgaver som påvirker konsistensen negativt, og de har en negativ verdi for diskrimineringssevne. Alle oppgavene i K1 følger samme struktur, hvor elevene fikk informasjon som var relevant for økosystemene på havbunnen utenfor Alaska, og skulle vurdere om en rekke forklaringer var mulige eller ikke ut fra informasjonen som var gitt. Etter revideringen fra studie 3 står det eksplisitt «*En forklaring er mulig, hvis den ikke utelukkes fra informasjonen du får.*» i forkant av oppgavene. Oppgavene ble gitt i to omganger (del 2.2 og 2.3), slik at elevene vurderte samme forklaring to ganger, men det ble gitt mer informasjon om casen i del 2.3 enn i 2.2, og flere forklaringer kunne utelukkes. Det kan derfor virke som at mange av elevene ikke fikk med seg at alle forklaringer skal anses som mulige, hvis de ikke utelukkes av informasjonen som blir gitt. Den forklaringen som er den korrekte i casen, møter elevene først i oppgave 2.2 d) ( $d = -0,20$ ,  $p = 0,38$ ) og senere i 2.3 e) ( $d = -0,10$ ,  $p = 0,47$ ), og denne forklaringen utelukkes ikke av informasjonen som elevene får. Likevel velger flertallet av elevene å svare at «*Dette er ikke en mulig forklaring*». Den negative diskrimineringssevnen til de to oppgavene viser at både de som scorer høyt i delferdigheten eller til kategorien scorer dårligere enn gjennomsnittet på de to oppgavene. Det kan virke som at elevene utelukker forklaringen fordi de synes den virker usannsynlig, slik som to naturfagslærere gjorde i studie 2. Delferdigheten blir indirekte målt gjennom oppgavene i K3. Delferdigheten K3 undersøker om elevene kom frem til riktig konklusjon etter å ha fått tilgang til all informasjonen tilknyttet casen. Å fjerne alle de 12 oppgavene tilknyttet delferdighet K1 er et mulig alternativ, som vil redusere omfanget av kartleggingsverktøyet, og gjøre oppgavefordelingen likere mellom de tre kategoriene i testen. Hvis det er hensiktsmessig å beholde oppgavene i K1, bør stammen i oppgavene endres. En alternativ formulering av oppgavestammen bør fokusere på om forklaringen har blitt utelukket av informasjonen som er gitt, i stedet for å spørre om den er «mulig», ettersom dette gir respondentene et større tolkningsrom.

Den interne konsistens ( $\alpha = 0,27$ ) til delferdigheten «Vurdere bevisgrunnlag» (K2) er under akseptabelt nivå. Ingen av oppgavene diskriminerer på et akseptabelt nivå i forhold til delferdigheten, selv om Tabell 4-17 viser at 2.1 c) diskriminerer over akseptabelt nivå i forhold til testkategorien. At delferdigheten måles gjennom kun fire oppgaver, bidrar negativt til den interne konsistensen i delferdigheten. Ettersom tre av oppgavene tilknyttet delferdigheten K2 i testen er henger sammen med oppgavene i K1, og ikke diskriminerer på akseptabelt nivå, bør også denne delferdigheten ekskluderes, hvis K1 fjernes fra testen. Elevene må vurdere bevisgrunnlaget for å klare oppgavene i del 2 av testen, og delferdigheten

måles dermed indirekte gjennom for eksempel oppgave 2.4 b) og 2.5 c). I tillegg gjør elevene rede for hvordan de brukte tilgjengelig informasjon for å finne riktig forklaring på mysteriet i casen i oppgave 2.4 c), som er en åpen oppgave tilknyttet S1. Oppgave 2.1 c) er en åpen oppgave som diskriminerer over akseptabelt nivå i forhold til kategorien, og kan for eksempel inkluderes i delferdigheten K3, hvor den diskriminerer på et akseptabelt nivå ( $d=0,4$ ).

Delferdigheten «konkludere» (K3) måler elevenes evne til å trekke en konklusjon, og har en akseptabel intern konsistens ( $\alpha=0,66$ ), selv om det kun er fire oppgaver tilknyttet delferdigheten. I oppgave 2.4 b) har elevene blitt gitt all nødvendig informasjon for å utelukke alle forklaringene utenom forklaringen: «*Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggere har endret matvaner fra fisk til sjøotere*». Denne oppgaven er blant de oppgavene i testen som korrelerer sterkest i forhold til score på ferdighetstesten ( $r=0,57$ ). Etter å ha blitt presentert for mer informasjon om økosystemene, gjennom en graf, spør oppgave 2.5 c) samme spørsmål som oppgave 2.4 b), også denne oppgaven korrelerer tydelig med poengsum på ferdighetstesten ( $r=0,54$ ). Det er altså de samme elevene som klarer å utelukke og resonere seg frem til riktig forklaring på 2.4 b) og 2.5 c), som viser høy kompetanse i testen. Oppgave 1.3 b) og 1.3 c) diskriminerer svakere enn akseptabelt nivå, og bidrar negativt til den interne konsistensen i delferdigheten. Årsaken til forskjellen mellom oppgavene kan være at 2.4 b) og 2.5 c) tilhører del 2 av testen, og har fem distraktorer, og er vanskeligere å gjette på, enn 1.3 b) og 1.3 c) tilhører del 1, som er av oppgavetyper AC, som altså er enklere å gjette på. Det kan tyde på at oppgavetype og hvilken del av testen oppgaven tilhører påvirker intern konsistens og diskrimineringssevne mer enn hvilken delferdighet den tilhører.

Som vist til i kapittel 4.3 ble det gjort endringer i testen etter studie 3. En av endringene var introduksjonen av delferdigheten «vurdere argumenter» (K4) som del av kategorien «Trekke velbegrunnede konklusjoner». Delferdigheten er en forenkling av den tidligere testkategorien «vurdere ut fra kriterier». Den interne konsistensen for delferdigheten ( $\alpha=0,45$ ) er under akseptabelt nivå, og kun to av oppgavene har akseptabel diskrimineringssevne. Oppgave 1.3 g) påvirker den interne konsistensen negativt, mens de resterende oppgavene påvirker den interne konsistensen positivt. Aktuelle endringer kan være å ekskludere eller erstatte oppgave 1.3 g), og revidere resten av oppgavene. Dette kan gjøres ved å fremheve at elevene skal vurdere relevansen argumentene har til spørsmålet, og ha en begrepsforklaring for «*relevans*».

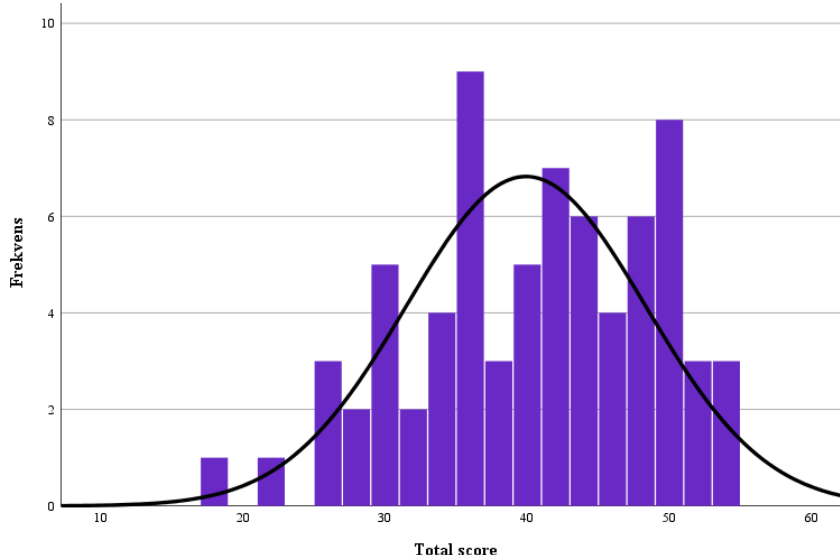
### Hva forteller resultatene om testen som helhet?

Hvor godt mål score på ferdighetstesten er for å måle de ferdighetene som er nødvendige for kritisk tenkning i naturfag drøftes i kapittel 5. I denne delen av oppgaven vil resultatene fra de statistiske analysene fremlegges. Tabell 4-18 viser at det er intern konsistens i ferdighetstesten som helhet ( $\alpha=0,86$ ), som er innenfor akseptabelt nivå ( $\alpha>0,65$ ). Standardfeilen for hele testen er under ett poeng, og indikerer at scorene til enkeltrespondenter er pålitelige. Den gjennomsnittlige vanskelighetsgraden ( $p=0,69$ ) for hele testen er høyere enn ønskelig ( $p=0,50$ ). Det vil si at testen er enklere enn den optimale vanskelighetsgraden, med hensyn til å differensiere ulike nivåer av kompetanse blant elevene.

Tabell 4-18: Resultater om ferdighetstesten som helhet

<b>Hel test</b>						
<b>Cronbach alfa</b>	<b>Vanskelighetsgrad (p)</b>	<b>Gj. Snitt i poeng</b>	<b>Standardavvik</b>	<b>Standardfeil</b>	<b>Antall oppgaver</b>	<b>Antall respondenter (n)</b>
0,86	0,69	39,89	8,45	0,99	53	72

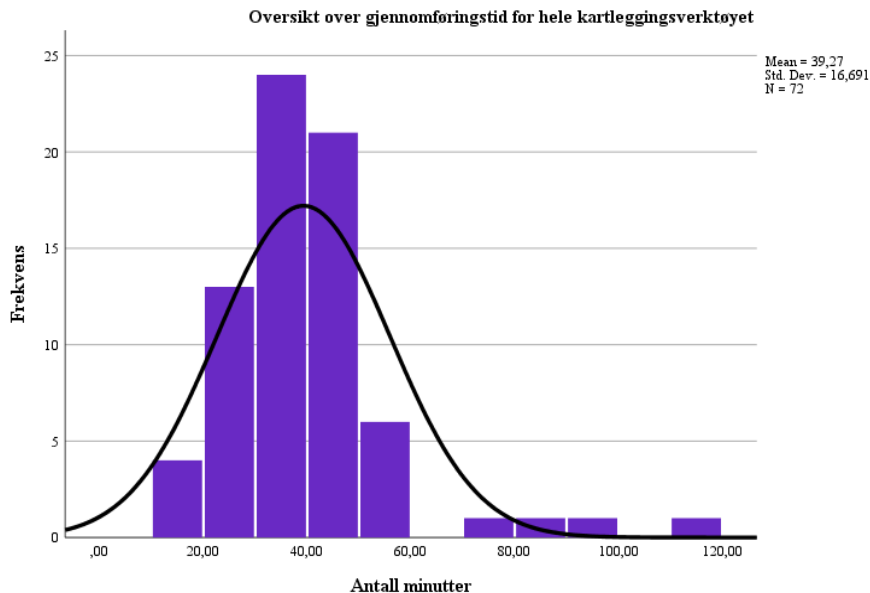
Figur 4-1 viser fordelingen av sammenlagt score på ferdighetstesten, og at det ikke er en perfekt normalfordeling av utvalgets scorer. Analysen av Kolmogorov-Smirnov-test viste en signifikantverdi på 0,20, som indikerer at resultatene for utvalg 4 ikke er normalfordelt, og ferdighetstesten ikke kan anses som en parametriske test med utgangspunkt i resultatene fra utvalget.



Figur 4-1: Stolpediagram som viser fordeling av resultater på ferdighetstest.

For at en test skal være anvendelig må den være praktisk gjennomførbar. Figur 4.2 viser at elevene gjennomsnittlig brukte 39,27 minutter på hele kartleggingsverktøyet. De elevene som

brukte mer enn 60 minutter fikk pause, og ifølge lærerne gjennomførte elevene testen når de fikk tid i timen etter. De arbeidet altså ikke kontinuerlig med kartleggingsverktøyet, og dette påvirker gjennomsnittstiden. Den oppgitte gjennomsnittstiden blir altså lengere enn reell gjennomsnittstid. Medianen (37 minutter) til gjennomføringstiden kan derfor være et bedre mål å ta utgangspunkt i enn gjennomsnittet.



Figur 4-2: Stolpediagram over gjennomføringstid for hele kartleggingsverktøyet.

Det ble undersøkt om resultatene på delferdighetene samsvarte med testscore. Dette kan gi svar på om det er de samme elevene som scorer høyt i de ulike delferdighetene som scorer høyt generelt i testen. Dette kan minne om hvordan oppgavers diskrimineringssevne analyserer, men det er ikke nødvendigvis de samme elevene som scorer høyt på de ulike delferdighetene. Tabell 4-19 viser at det er sterke korrelasjoner mellom de enkelte delferdighetene og testen som helhet. Delferdigheten K4, som ble introdusert etter revideringen av kartleggingsverktøyet i studie 3, er den eneste kategorien som ikke har en tydelig korrelasjon med total testscore ( $r < 0,5$ ), men korrelasjonen er sterk nok til at K4 bidrar positivt til den interne konsistensen for hele testen. Resultatet viser at det er samsvar mellom delferdighetene i testen, med unntak av K4.

Tabell 4-19: Delferdighetenes påvirkning av testens reliabilitet og evne til å diskriminere

<b>Delferdighetenes påvirkning av testens reliabilitet og evne til å diskriminere</b>			
<b>Delferdighet</b>	<b>Vanskelighets-grad</b>	<b>Korrelasjon i forhold til hel test (r)</b>	<b>Cronbach's Alfa i ferdighetstesten, hvis delferdigheten slettes.</b>
S1	0,59	0,61	0,83
S2	0,84	0,65	0,83
N1	0,73	0,63	0,82
N2	0,75	0,67	0,82
K1	0,58	0,64	0,82
K2	0,65	0,58	0,83
K3	0,69	0,62	0,83
K4	0,60	0,41	0,85

#### *Reliabilitet i holdningskategoriene*

I kartleggingsverktøyet blir elevenes holdninger knyttet til kritisk tenking målt gjennom et spørreskjema. Holdningskategoriene anvendes for å undersøke mulige sammenhenger mellom testresultater og holdninger til kritisk tenkning og naturfag. Det er derfor viktig å få reliable resultater for at videre analyser skal gi troverdige resultater. Sammenlagt gir holdningene (Intellektuell nysgjerrighet, Åpenhet, Selvtillit, Sannhetssøkende og Bentenksomhet), som er basert på de holdningskategoriene som Facione et al. (2000) anvender, en akseptabel intern konsistens ( $\alpha=0,74$ ). Dette gir et reliabelt grunnlag for å undersøke sammenhenger mellom holdninger om kritisk tenkning generelt og andre variabler.

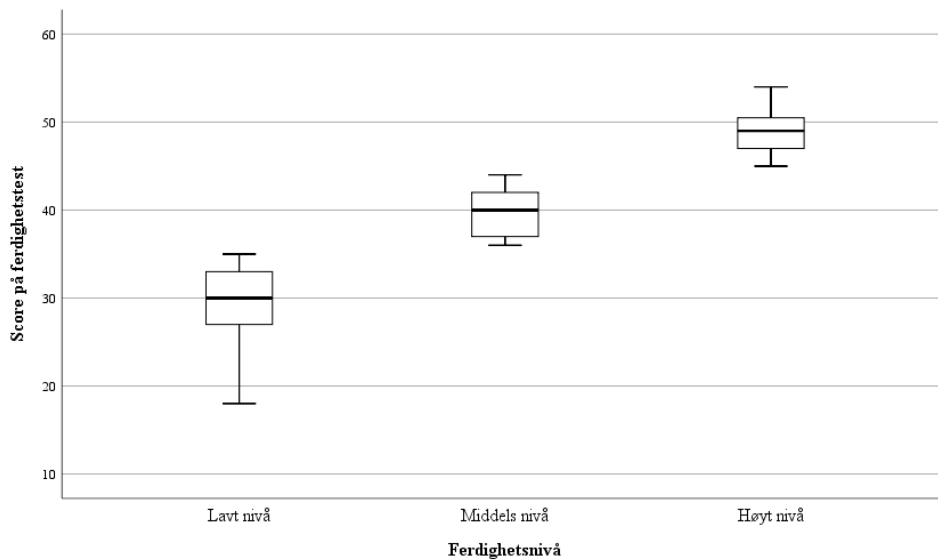
Tabell 4-20 viser en oversikt over Cronbach alfa for holdningskategoriene. Enkeltvis gir kun kategorien «åpenhet» akseptabel intern konsistens ( $\alpha=0,66$ ), mens de andre nevnte holdningskategoriene om kritisk tenkning har lav intern konsistens. Det er vedlagt en oversikt over resultatene knyttet til variablene i spørreskjemaet (se vedlegg 4.5). Enkelte variabler påvirket den interne konsistensen til holdningskategoriene negativt, og det er aktuelt å ekskludere slike holdningsvariabler (f.eks. I-2, A-4, Se-4, Sa-3 og Be-1) for å oppnå mer konsistente målinger i kategoriene. For eksempel ville den interne konsistensen til holdningskategorien «intellektuell nysgjerrighet» økt til akseptabelt nivå uten variabelen I-2. Holdningskategoriene som er spesifikke til naturfag og skolen gir mer reliable resultater. Alle variablene bidrar positivt til kategoriene «holdninger til naturfag» ( $\alpha=0,70$ ) og «Holdninger til viktigheten av kritisk tenkning i skolen» ( $\alpha=0,68$ ). Begge holdningskategoriene kan derfor gi pålitelige resultater for videre analyser.

	<b>Cronbach's alfa</b>	<b>n spørsmål</b>
<b>Kritiske holdninger (I, A, Se, Sa og B sammenlagt)</b>	0,74	20
<b>Holdningskategori</b>	<b>Cronbach's alfa</b>	<b>n spørsmål</b>
Intellektuell nysgjerrighet (I)	0,52	4
Åpenhet for ny informasjon (A)	0,66	4
Selvtillit (Se)	0,38	4
Sannhetssøkende (Sa)	0,30	4
Betenksomhet ovenfor avgjørelser (B)	0,38	4
Holdninger til <b>naturfag</b> (N)	0,70	4
Holdninger til viktigheten av <b>kritisk tenkning</b> i skolen (H)	0,68	3

Tabell 4-20: Resultater til intern konsistens til holdningskategoriene

#### 4.4.3 Resultater knyttet til Forskningsspørsmål 2: Testens resultater

Utenom ferdighetstesten er formålet med variablene i kartleggingsverktøyet å samle inn informasjon om utvalget, og undersøke hvilke variabler som korrelerer med score i testen. Forskningsspørsmål 2, «Hva kjennetegner testens resultater?», omhandler alle resultater som kan fortelle noe om respondentene. Spesielt fokus tilegnes delspørsmålene «Hvordan er vanskelighetsgraden til de ulike delferdighetene og testen som helhet?» (FS2.a) og «Hvordan skiller testen mellom ulike elevgrupper» (FS2.b). For å undersøke testens resultater har utvalget blitt delt inn i tre grupper. Dette gjør det mulig å undersøke om det er samsvar mellom elever som scorer høyt og lavt i de ulike testkategoriene og testen for øvrig. Gruppene tar utgangspunkt i elevens samlede score på ferdighetstesten. Med utgangspunkt i 33. og 66. percentile ble elever som scorer 35 poeng eller lavere, samlet i en gruppe som videre vil tiltales «lavt ferdighetsnivå». De med «høyt ferdighetsnivå» scoret 45 poeng eller høyere, og de med «middels ferdighetsnivå» hadde en score mellom 36 og 44 poeng. Det tas forbehold om at testen kun gir en observert score på elevenes ferdigheter, og hvor mye ferdighetsnivåene forteller oss om elevenes sanne kompetanse til å tenke kritisk er avhengig av testens kvalitet, som diskuteres i kapittel 5. Variasjonen av testscorer innenfor i de ulike ferdighetsnivåene vises i Figur 4-3. Det er størst variasjon blant elevene på lavt ferdighetsnivå, med enkeltelever som scorer opptil 10 poeng unna medianen. Kvartilbredden er minst for elevene på høyt ferdighetsnivå, som vil si at scorene til elevene i denne gruppen er mest samlet.



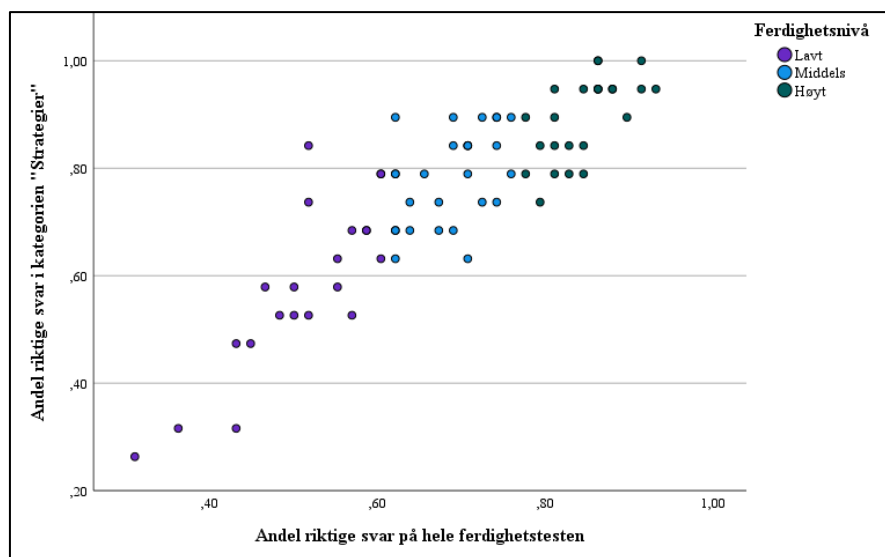
Figur 4-3: Boksdiagram med oversikt over variasjon i score på ferdighetstesten for de tre ferdighetsnivåene.

#### Resultater knyttet til forskningsspørsmål 2.a.

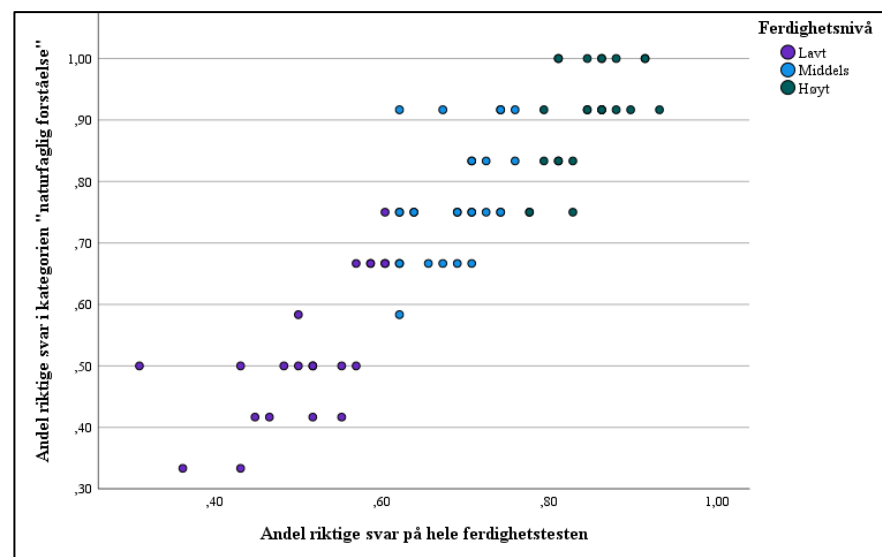
Det er interessant å undersøke om det er de samme elevene som får til ulike elementer i testen. I så fall bør elevens ferdighetsnivå samsvare med vanskelighetsgraden på testens kategorier og delferdigheter. Figur 4-5, 4-6 og 4,7 viser fordelingen av elevenes score på de tre kategoriene i testen, opp mot elevenes sammenlagte poengsum. For å gjøre sammenligning enklere har score i testen og i testkategori blitt omgjort til prosent riktig. Blant testkategoriene er det score i kategorien «konklusjoner» som korrelerer sterkest med score i ferdighetstesten ( $r=0,91$ ,  $p<0,01$ ), men dette er ikke overaskende med tanke på at det er den mest omfattende kategorien, med hensyn til antall oppgaver. Det er også tydelig korrelasjon mellom testscore og score på de to resterende testkategoriene, «strategier for kritisk tenkning» ( $r=0,87$ ,  $p<0,01$ ) og «naturfaglig forståelse» ( $r=0,90$ ,  $p<0,01$ ). Det er altså i stor grad de samme elevene som får til oppgavene i de ulike kategoriene. Den kategorien med lavest vanskelighetsgrad var «strategier for kritisk tenkning» ( $p=0,76$ ), og som vist i Figur 4-7 fikk elever på middels og høyt ferdighetsnivå til de fleste oppgavene i denne kategorien, mens variasjonen er større blant elever på lavt nivå. Kategorien «Naturfaglig forståelse» er noe mer utfordrende ( $p=0,74$ ), mens kategorien «konklusjoner» er den testkategorien med høyest gjennomsnittlig vanskelighetsgrad ( $p=0,61$ ). Spredningsplottet i Figur 4-6 viser at det kun er elever på høyt ferdighetsnivå som fikk til over 80% av oppgavene innad i kategorien «konklusjoner». Det er også interessant at det er en tydelig korrelasjon ( $r=0,74$ ,  $p<0,01$ ) mellom hvilke elever som fikk høy score på de åpne oppgavene i testen og som fikk høy score i testen generelt. Likevel



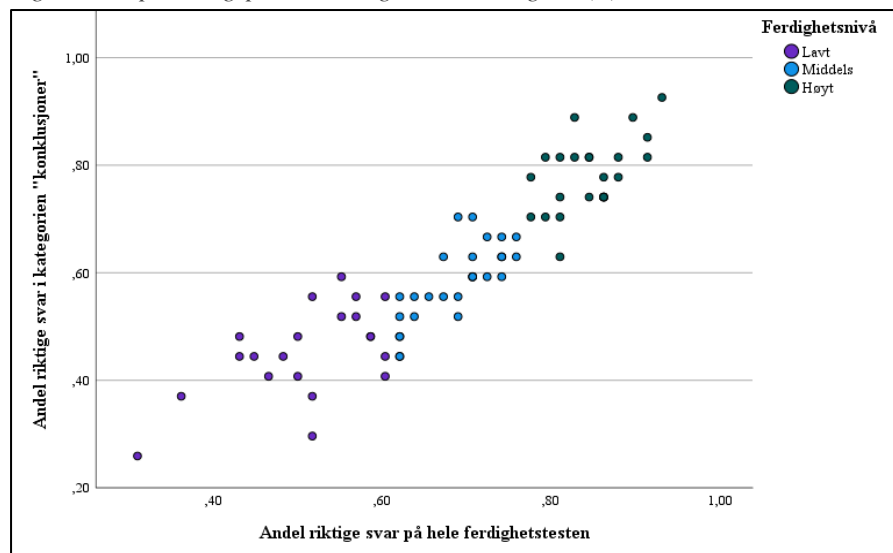
viser Figur 4-4 at det er større spredning i testscoren til elevene som scoret høyt på de åpne oppgavene. Dette betyr at elever på høyt ferdighetsnivå gjennomsnittlig scoret høyere enn elever på lavt ferdighetsnivå, men som man kan se i Figur 4-4 er det ikke et klart skille mellom de tre ferdighetsnivåene.



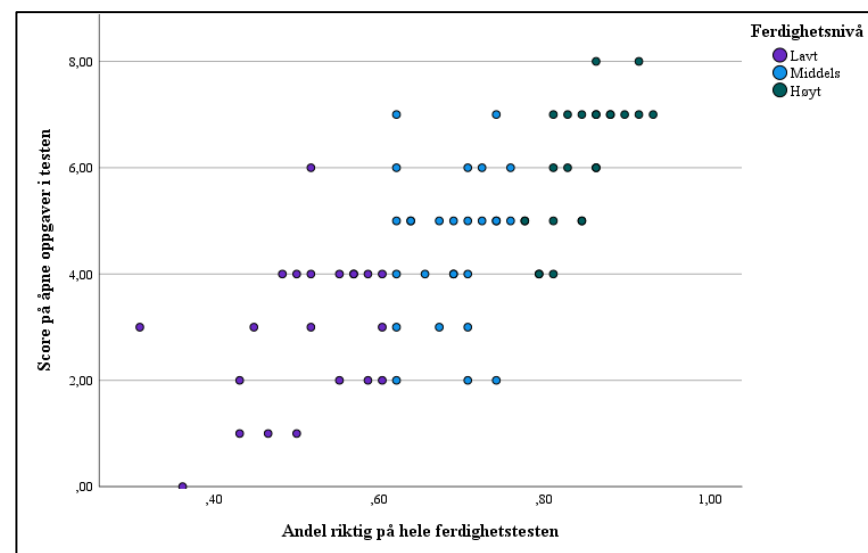
Figur 4-7: Spredningsplott av kategorien "strategier" (S)



Figur 4-5: Spredningsplott av kategorien "naturfaglig forståelse" (N)

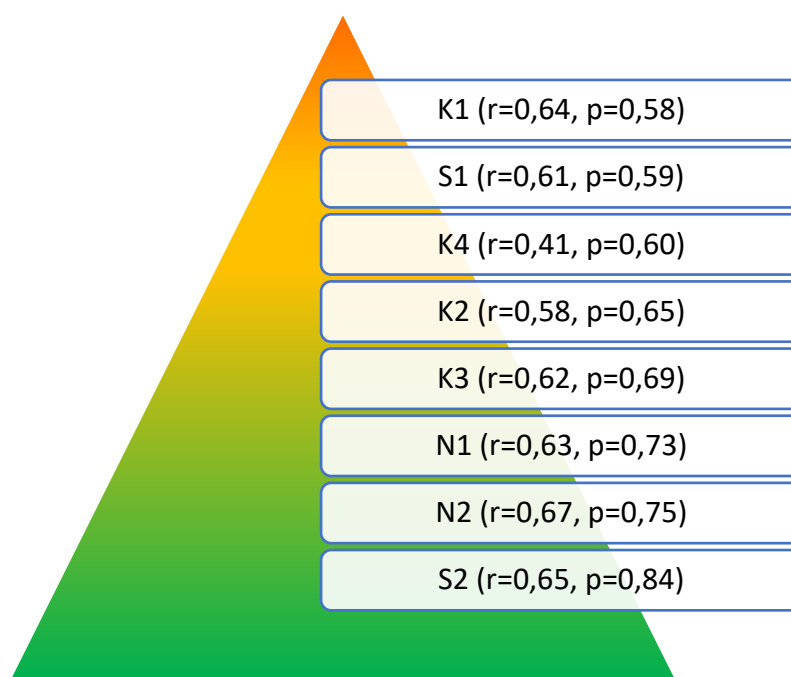


Figur 4-6: Spredningsplott med kategorien "konklusjoner" (K)



Figur 4-4: Spredningsplott av score på de åpne oppgavene i testen

Også på delferdighetsnivå ble korrelasjon til testscore undersøkt, for å få et bedre overblikk over vanskelighetsgraden til de ulike delferdighetene. Delferdigheter som korrelerer tydelig ( $r > 0,5$ ) med score på ferdighetstesten kan være indikatorer for hva som er enkelt og vanskelig for elever på de ulike ferdighetsnivåene. Alle delferdighetene korrelerer tydelig med testscore, med unntak av delferdigheten «vurdere argumenter» (K4) ( $r=0,41$ ). Etersom det er samsvar mellom delferdighetene og ferdighetsnivå kan vanskelighetsgraden være en god indikator for hvilke delferdigheter som er utfordrende og enkle for elever på ulike ferdighetsnivåer. Figur 4-8 gir en oversikt over vanskelighetsgraden til de ulike delferdighetene og hvor sterkt delferdigheten korrelerer med testscore.

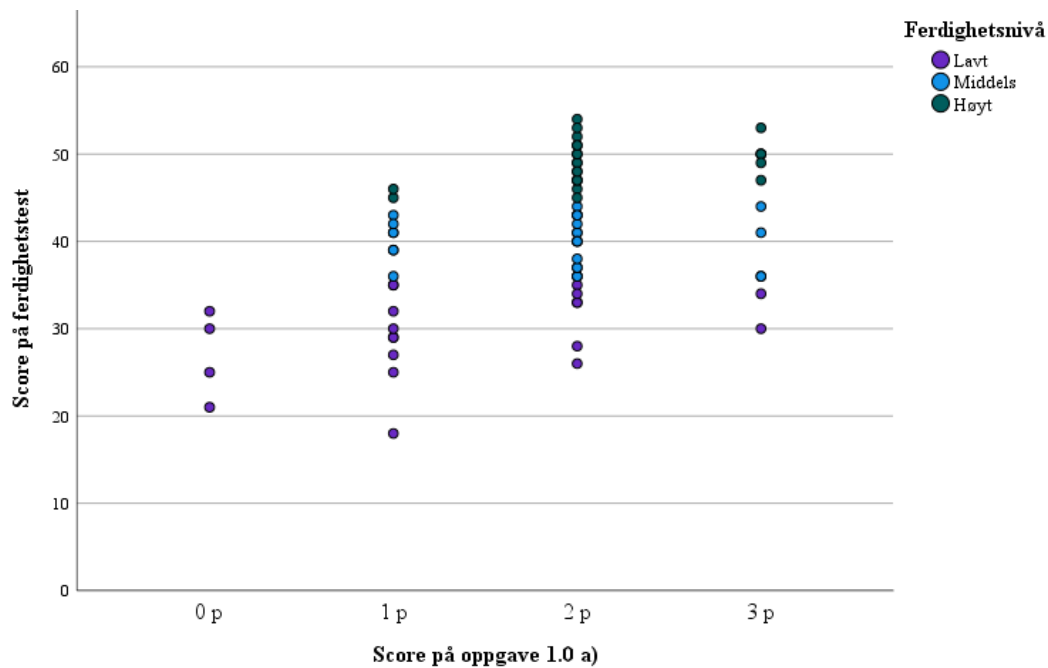


Figur 4-8: Rangering vanskelighetsgraden til delferdighetene

I Figur 4-8 kan vi se at de fleste elevene på alle ferdighetsnivåer klarte i stor grad å «vurdere troverdigheten» (S2) til de to tekstene i testens del 1 ( $p=0,84$ ), og delferdigheten ble i stor grad målt gjennom AC-oppgaver med påstander om tekstenes troverdighet. Begge delferdighetene innen kategorien som måler «forståelse for naturfaglig informasjon» (N) hadde en lav vanskelighetsgrad. Dette tyder på at de fleste elevene i stor grad klarte å «avkode og tydeliggjøre informasjon fra tekster» (N1) ( $p=0,73$ ) og «Avkode informasjon fra tabeller og illustrasjoner» (N2) ( $p=0,75$ ). Delferdighetene under kategorien «konklusjoner» (K) viste seg å være mer utfordrende for utvalget, og differensierer dermed bedre mellom ferdighetsnivåene.

Delferdigheten som måler elevenes evne til å ta velbegrunnede konklusjoner (K3) har en sammenlagt p-verdi på 0,69. I de fire oppgavene som er tilknyttet delferdigheten, er det stor variasjon i vanskelighetsgrad. Oppgave 1.3 b) ( $p=0,90$ ) og 1.3 c) ( $p=0,89$ ) er knyttet til del 1 og nesten alle elevene fikk poeng på oppgavene. De to oppgavene er ikke egnet for å undersøke forskjellen på ferdighetsnivåene, fordi vanskelighetsgraden er for lav. Oppgavene som undersøker om elevene klarte å finne den mest sannsynlige forklaringen på casen i del 2, altså oppgave 2.4 b) ( $p=0,45$ ) og 2.5 c) ( $p=0,49$ ) var det nesten ingen elever på lavt ferdighetsnivå som klarte, mens de fleste av de på høyt ferdighetsnivå fant nøkkelen i oppgavene. Å trekke en velbegrunnet og logisk konklusjon basert på det omfattende informasjonsgrunnlaget elevene måtte forholde seg til i del 2 var altså utfordrende for mange elever. Delferdigheten «vurdere bevisgrunnlag» (K2) ( $p=0,65$ ), som ble foreslått å fjerne tidligere i kapitlet, var noe mer utfordrende enn K3.

De tre delferdighetene som har høyest vanskelighetsgrad har en p-verdi mellom 0,60-0,58, og bør være egnet for å undersøke hva som gjennomsnittlig var mer utfordrende for elever med middels ferdighetsnivå, sammenlignet med elever på høyt ferdighetsnivå. Delferdigheten «Vurdere argumenter» (K4) ( $p=0,60$ ) var utfordrende for elevene, men det er ikke en tydelig korrelasjon mellom score på hele ferdighetstesten og delferdigheten ( $r=0,4$ ). Det vil si at oppgavene i delferdigheten ikke diskriminerer mellom elever på lavt, middels og høyt ferdighetsnivå på samme måte som de andre delferdighetene. Dette var den delferdigheten som elevene på høyt ferdighetsnivå gjennomsnittlig scoret svakest på. Evnen til å forklare, begrunne og reflektere over egne strategier måles gjennom delferdigheten S1, og kjennetegnes av en p-verdi på 0,59. S1 består av kun to åpne oppgaver, hvor elevene skal gjøre rede for hvilke strategier anvender. Oppgave 1.0 a) er en av de to oppgavene knyttet til delferdigheten, og er et eksempel på hvordan poengene kan fordele seg på en åpen oppgave. Som vist i Figur 4-9, var elever på høyt ferdighetsnivå gjennomsnittlig bedre på å gjøre rede for egne strategier knyttet til kildekritikk enn elever på lavt ferdighetsnivå.



Figur 4-9: Spredningsplott av score på oppgave 1.0 a) og score på ferdighetstest, inndelt etter ferdighetsnivå

Delferdigheten som undersøker elevenes evne til å «undersøke og vurdere bevis» (K1) er den delferdigheten med lavest vanskelighetsgrad ( $p=0,58$ ). Elevenes score i delferdigheten korrelerer tydelig ( $r=0,64$ ) med elevenes score på hele ferdighetstesten, som tyder på at elever på høyt ferdighetsnivå gjennomsnittlig gjorde det bedre enn elever på lavere ferdighetsnivå. K1 er delferdigheten med minst intern konsistens blant de tilknyttede oppgavene. Resultatene som er presentert over viser vanskelighetsgraden til delferdighetene i testen. I kapittel 5.2 vil det drøftes i hvor stor grad resultatene kan fortelle oss noe om hvilke ferdigheter som er mest utfordrende for elevene i virkeligheten.

Det ble også gjennomført en klasseromsamtale med en av klassene i utvalg 4, hvor jeg spurte hva elevene synes var enkelt og utfordrende i testen. Flere av de elevene som uttrykket seg i samtalen mente at del 1 av testen var enklere enn del 2. Noen av elevene uttrykket at oppgavene i del 2 var interessante og morsomme å jobbe med, og at de måtte tenke mye for å finne løsningen på casen. En elev fortalte at han hadde begynt å tegne grafer selv i løpet av del 2.2 og 2.3 for å lage en oversikt over hvilke arter som det ble flere og færre av i økosystemet, og laget et tankekart med oversikt over hvordan de ulike artene påvirker hverandre. Ifølge han selv førte dette han til den riktige konklusjonen.

### Resultater knyttet til forskningsspørsmål 2.b

Ulike hypotesetester ble anvendt for å undersøke hvor godt ferdighetstesten skiller mellom ulike grupper av elever (FS2.b). Det er i utgangspunktet ikke utformet hypoteser på forhånd om hvordan ulike elevgrupper vil score på testen. Informasjon om elevene ble innhentet via bakgrunnsvariablene i starten av kartleggingsverktøyet. Tabell 4-21 viser resultatene fra uavhengige t-tester som undersøkte score på ferdighetstesten og ulike variabler. Som nevnt i kapittel 3 kan ikke resultatene generaliseres utover utvalget, selv ved signifikante forskjeller.

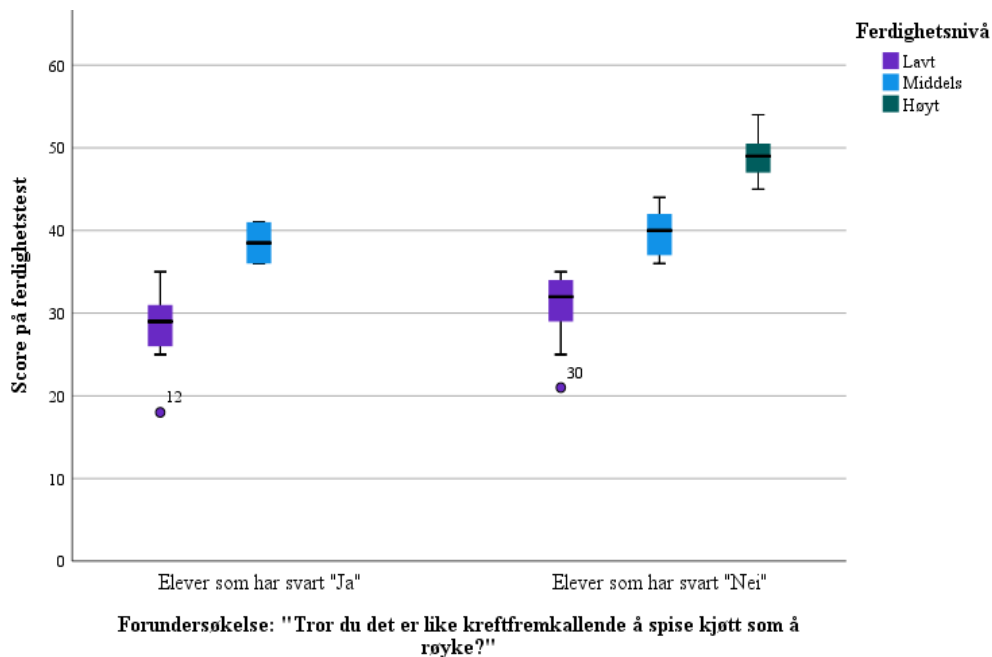
Tabell 4-21: Resultater på T-tester om elevgrupper

	Svarsalternativ (n)	Forskjell i gj. Snitt testscore	Sig. (p)	Effekt (Cohens's d)
Trinn	10. trinn (33) 9. trinn (39)	3,03 p	0,13	0,36
Kjønn	Gutt (31) Jente (40) («annet» (1) ikke med i t-test)	0,99 p	0,20	0,12
Minoritetsspråklig bakgrunn	Minst en nordisk forelder (46) Ikke nordiske foreldre (26)	8,04 p	<0,01	1,07
Karakter naturfag	«5 eller 6» (38) «3 eller 4» (36) («1 eller 2» (2) ikke med i t-test)	6,55 p	<0,01	0,83
Karakter norsk	«5 eller 6» (44) 3 eller 4 (28) («1 eller 2» (0) ikke med)	6,18 p	<0,01	0,78
Forundersøkelse 1:	Ja (40) Nei (32)	0,78 p	0,70	0,09
Forundersøkelse 2:	Ja (10) Nei (62)	11,27 p	<0,01	1,50

Elever fra 10. trinn scoret i gjennomsnitt 3,03 poeng bedre enn elever fra 9.trinn. Dette er ikke en stor nok forskjell for å gi signifikante forskjeller ( $p=0,13$ ,  $d=0,36$ ) i utvalget, ettersom signifikansnivået som anvendes for studien er satt til 5% ( $p<0,05$ ). Jentene i utvalget fikk gjennomsnittlig 1 poeng mindre enn guttene, men forskjellen er ikke signifikant i utvalget ( $p=0,20$ ,  $d=0,12$ ). Det var betydelige forskjeller mellom elever som har minst en nordisk forelder og de som ikke har det. Elever uten minoritetsspråklig bakgrunn scoret i gjennomsnitt 8,04 poeng høyere enn elever med minst en minoritetsspråklig forelder. For utvalget gir dette signifikante forskjeller ( $p<0,01$ ,  $d=1,07$ ). Høye karakterer i naturfag ( $p<0,01$ ,  $d=0,83$ ) og norsk ( $p<0,01$ ,  $d=0,78$ ) virker til å være betydningsfullt for hvor høy score elevene fikk på ferdighetstesten. Forskjellene i gjennomsnitt mellom de to variablene var over 6 poeng.

I spørreskjemaet som elevene skulle fylle ut før testen, fikk elevene to spørsmål knyttet til kreftrisiko ved å spise kjøtt, som er navngitt «forundersøkelse 1» og «forundersøkelse 2» i Tabell 4-21. Utvalget var delt i om de mente at «*det kan være kreftfremkallende å spise kjøtt*»,

og forskjellen i gjennomsnittscore på ferdighetstesten var under ett poeng ( $p=0,70$ ,  $d=0,09$ ). Det var ingen signifikante forskjeller mellom hva elevene svarte på forundersøkelse 1. Det andre spørsmålet spurte om elevene tror «*det er like kreftfremkallende å spise kjøtt som å røyke*». Ti elever svarte ja, mens resten av utvalget svarte nei. De som svarte ja fikk i gjennomsnitt 11,27 færre poeng enn de som svarte nei. Forskjellen er signifikant ( $p<0,01$ ,  $d=1,5$ ). Figur 4-10 viser at ingen elever på høyt ferdighetsnivå svarte at de trodde det var like kreftfremkallende å spise kjøtt som å røyke.



Figur 4-10: Spredningsplott av score på ferdighetstest og svar på forundersøkelse om kreftfaren ved røyking.

Hvor mye elevene leser på fritiden og leseforståelse ble forsøkt undersøkt med to variabler, og om dette har sammenheng med testscore ble analysert med en-veis-ANOVA test. Elevene ble spurt «*hvor mye leser du på fritiden?*». Elevene som svarer at de «*leser mye*» ( $n=13$ ) scoret best på testen (gj. score= $41,46$ ), de som «*leste litt*» ( $n=30$ ) scoret gjennomsnittlig 1,23 poeng mindre, mens de som svarte at bare leser når de «*må*» ( $n=29$ ) scoret 2,6 poeng lavere enn de som leste mye. Forskjellen mellom de tre gruppene er ikke signifikant ( $p=0,63$ ).

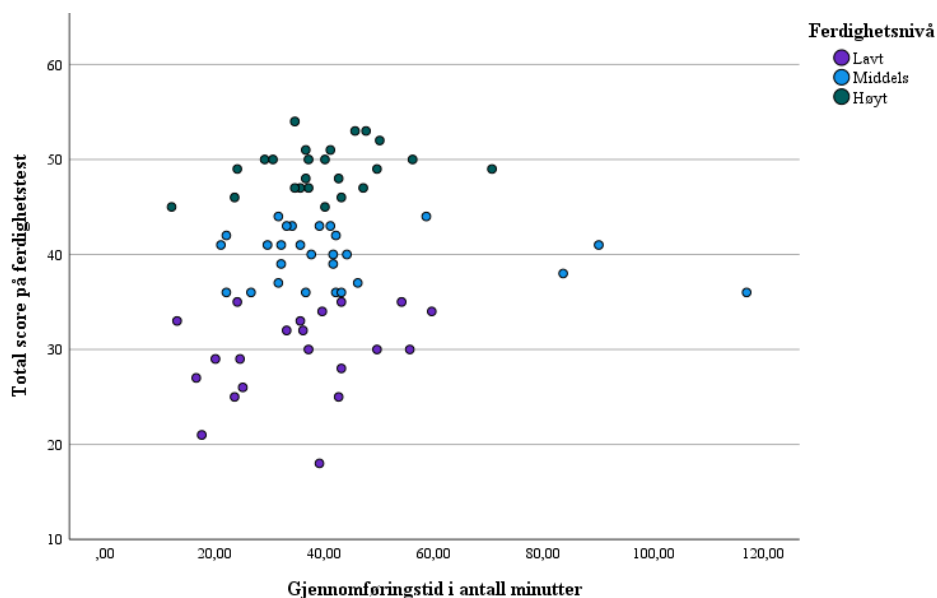
Leseforståelse ble målt gjennom en variabel som spør hvor mye elevene forsto av det de leste forrige gang de leste i en lærebok i naturfag. Tabell 4-22 gir en oversikt over resultatene fra analysen i ANOVA, og viser at det ikke er signifikante forskjeller i gjennomsnittet for elever som valgte de ulike svarsalternativene ( $p=0,36$ ). Det er interessant å se at elevene som svarte at de «*forsto alt*» de leste faktisk fikk gjennomsnittlig lavere poengsum enn de som «*forsto det meste*». Ettersom flertallet av elevene svarte at de enten «*forsto en del*» eller «*forsto det*

meste», ble det gjennomført en T-test på de to gruppene. T-testen viste at forskjellen mellom gruppene ikke er signifikante ( $p=0,09$   $d=0,47$ ).

Forståelse av faginnhold fra lærebok	N	Gj. Score på ferdighetstest
«Forsto lite»	2	35,00
«Forsto en del»	20	37,55
«Forsto det meste»	41	41,29
«Forsto alt»	9	39,89
Total	72	39,90

Tabell 4-22: Resultater på ANOVA med variabelen leseforståelse

Det er virker ikke til å være sammenheng mellom score på ferdighetstesten og gjennomføringstid. Utgangspunktet for påstanden er at Figur 4-11 viser stor spredning i gjennomføringstid blant elever på alle tre ferdighetsnivåene, og det er ikke signifikant korrelasjon mellom antall minutter gjennomføringstid og totalscore ( $r=0,13$ ,  $p=0,29$ ).

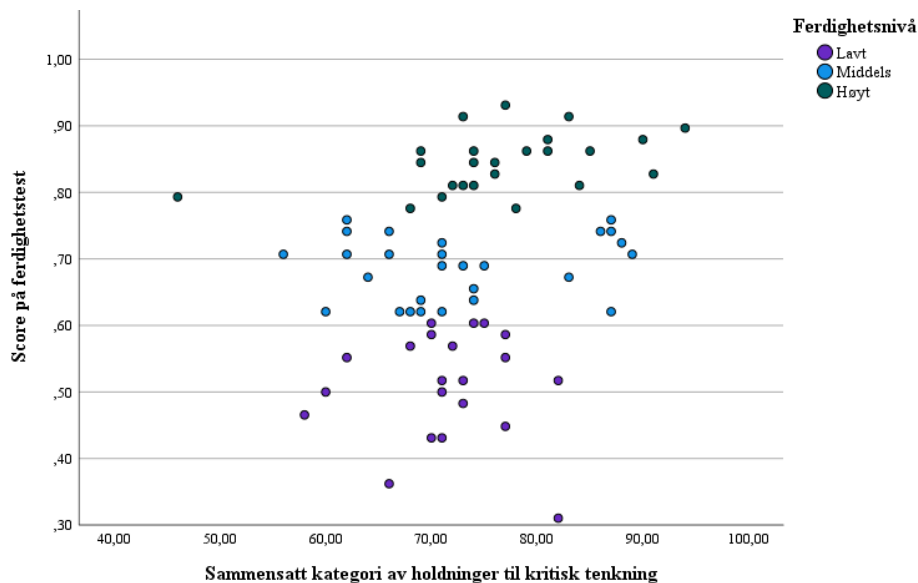


Figur 4-11: Spredningsplott av score på ferdighetstest og gjennomføringstid inndelt etter ferdighetsnivå.

Holdningskategoriene ble anvendt for å undersøke mulige sammenhenger mellom ferdigheter og holdninger til kritisk tenkning og naturfag. I spørreskjemaet ble fem holdningskategorier (Intellektuell nysgjerrighet, Åpenhet, Selvtillit, Sannhetssøkende og Bentensksomhet) som var basert på Facione et al. (2000) sine holdningskategorier knyttet til kritisk tenkning undersøkt, i tillegg til to som handlet om kritisk tenkning i skolen og holdninger knyttet til viktigheten av



naturfag. Sammenlagt gir de fem holdningskategoriene om kritisk tenkning en akseptabel intern konsistens ( $\alpha=0,74$ ). Dette gir et reliabelt grunnlag for å undersøke sammenhenger mellom holdninger om kritisk tenkning generelt og resultat på ferdighetstesten. For å undersøke mulige sammenhenger mellom holdninger og ferdigheter knyttet til kritisk tenkning ble det anvendt test for spearman's rho. Det er en liten korrelasjon ( $r_s=0,33$ ) mellom resultat på ferdighetstesten og positive holdninger til kritisk tenkning, og korrelasjonen er signifikant ( $p=0,01$ ). Korrelasjonen er illustrert i spredningsplottet i Figur 4-12.



Figur 4-12: Spredningsplott med score på ferdighetstesten og holdninger til kritisk tenkning

Korrelasjonstester av de enkelte holdningskategoriene innen kritisk tenkning viser at det er holdningskategorien «Sannhetssøkende» ( $r_s=0,39$ ,  $p<0,01$ ) som bidrar mest positivt til korrelasjonen mellom testscore og holdninger til kritisk tenkning. Det er verdt å nevne at denne holdningskategorien har lav intern konsistens ( $\alpha=0,30$ ). Holdninger knyttet til naturfag ble også målt, og korrelerer veldig svakt med resultater fra ferdighetstesten ( $r_s=0,09$ ,  $p=0,46$ ), men resultatene er ikke signifikante. Det samme gjelder for holdninger knyttet til viktigheten av kritisk tenkning i skolen ( $r_s=0,10$ ,  $p=0,42$ ).

## 5 Diskusjon

Hva forteller resultatene om kartleggingsverktøyet? Dette kapitlet har som formål å besvare dette spørsmålet med utgangspunkt i det teoretiske grunnlaget som ble fremlagt i kapittel 2. Hensikten med prosjektet er å utvikle og teste et måleinstrument for elevers kritiske tenkning i naturfag. Stegvis har ulike aspekter av kartleggingsverktøyet blitt undersøkt gjennom studie 1, 2 og 3. Mellom hver studie har det blitt gjennomført endringer for å øke kvaliteten til kartleggingsverktøyet, med hensyn til reliabilitet og validitet. Dette har resultert i kartleggingsverktøyet som ble anvendt i studie 4. En rød tråd gjennom oppgaven har vært operasjonaliseringen, utviklingen og kvalitetssikringen av ferdighetstesten, og det er denne delen av kartleggingsverktøyet som fortsatt vil være hovedfokuset gjennom dette kapitlet. Først drøftes testens reliabilitet, og videre gjøres det rede for hvordan jeg har forsøkt å ta hensyn til validitet i utformingen av testen som ble anvendt i studie 4. I den grad det er mulig vil jeg sammenligne resultatene fra ferdighetstesten med de eksisterende testene som ble gjort rede for i kapittel 2.4. Videre drøftes hva resultatene fra kartleggingsverktøyet kan fortelle om utvalget i studie 4, og hvilke begrensninger kartleggingsverktøyet og dette masterprosjektet har.

### 5.1 Hvor godt mål på elevers ferdigheter til å kritiske tenkning er ferdighetstesten?

Hensikten med ferdighetstesten er å måle kognitive ferdigheter som ungdomsskoleelever trenger for å kunne tenke kritisk i naturfag, og det optimale ville vært om elevenes score på testen ga direkte innsikt i elevenes faktiske ferdigheter. Ifølge klassisk testteori kan vi ikke ta utgangspunkt i at dette stemmer, for den scoren vi observerer gjennom elevenes besvarelser er summen av deres virkelige ferdigheter og den påvirkning som systematiske og tilfeldige feilkilder har på scoren. Testens kvalitet har derfor blitt gitt mye oppmerksomhet gjennom testutviklingen, for å minimere feilkildenes påvirkning på testscore. Under vil resultatene på kvalitetsmarkørene fra studie 4 anvendes som et utgangspunkt for å diskutere reliabiliteten ferdighetstesten, før validiteten drøftes.

#### 5.1.1 Reliabilitet

Det vanligste målet for å undersøke reliabiliteten til en test er gjennom å analysere dens interne konsistens. Som vist til i forrige kapittel er det intern konsistens ( $\alpha=0,86$ ) i ferdighetstesten som helhet, som består av 53 oppgaver. Dette på nivå med det kravet som settes til nasjonale prøver i regning ( $\alpha>0,85$ ) (Lie et al., 2005), og over det kravet som er satt i

dette masterprosjektet ( $\alpha > 0,65$ ). Til sammenligning ga testen i PROBE-prosjektet en Cronbach alfa på 0,84, basert på 39 oppgaver og et utvalg på 271 elever. Kluge et al. (2018) argumenterer for at det er den høyeste verdien som har blitt rapportert om innen testing av kritisk tenkning. Den anerkjente testen California Critical thinking test har en Cronbach alfa mellom 0,68 til 0,69 for de 34 oppgavene i testen (Facione, 1990b). Dette betyr at ferdighetstestens interne konsistens er på linje med eksisterende og anerkjente tester av kritisk tenkning. Den høye verdien av Cronbach Alfa i ferdighetstesten bør likevel ses i sammenheng med at testen består av flere oppgaver enn testene som beskrives i kapittel 2.4.

Verdien ( $\alpha = 0,86$ ) forteller oss at det er samsvar mellom hvilke respondenter som får mange eller få poeng på oppgavene i testen. Alle de tre kategoriene og de åtte delferdighetene bidrar positivt til testens interne konsistens. Det er tydelige korrelasjoner ( $r > 0,5$ ) mellom score på testen som helhet og de ulike kategoriene og delferdighetene, med unntak av en delferdighet (K4). Dette styrker testens evne til å anvendes som et reliabelt sammensatt mål på ferdigheter som muliggjør kritisk tenkning. Det er som nevnt ikke ønskelig med for høy intern konsistens i testen ( $\alpha > 90$ ) (Tavakol & Dennick, 2011), ettersom testen undersøker ulike kategorier og delferdigheter, og det er hensiktsmessig å ha variasjon i oppgavene, med tanke på validitet. En test som kun stiller det samme spørsmålet vil ha svært god intern konsistens, men vil ikke gi spesielt interessante resultater.

Den høye interne konsistensen gir rom for å redusere antall oppgaver, slik det foreslås i kapittel 4.4.2. Dette kan medføre at det blir lavere intern konsistens, ettersom antall testledd påvirker Cronbach Alfa. Men det er ikke sikkert, ettersom flere av de oppgavene som foreslås å ekskluderes bidrar negativt til testens eller testkategoriens interne konsistens. Det er ønskelig å gjøre kartleggingsverktøyet mindre omfattende, med hensyn til gjennomføringstid, elevenes utholdenhet og konsentrasjonsevne, som Cohen et al. (2007) kaller testens «*practicability*». Målet har vært at kartleggingsverktøyet skal være mulig å gjennomføre med en klasse i løpet av en klokke time, medberegnet tid til å introdusere kartleggingsverktøyet for elevene. Kun fire elever brukte mer enn seksti minutter, og medianen for gjennomføringstiden var 37 minutter. Dette er beregnet fra elevene startet på testen, og inkluderte ikke den tiden læreren brukte på å introdusere undersøkelsen. Testen består av 53 oppgaver, men kartleggingsverktøyet undersøker flere variabler enn ferdigheter, så undersøkelsen bør anses som omfattende. Antall oppgaver og spørsmål kan ha ført til at elever mistet motivasjonen eller konsentrasjonen underveis. Dette kan ha påvirket resultatene. Valget om å inkludere

mange oppgaver var bevisst, ettersom testen fortsatt er i en pilotfase. Dette gir mulighet til å eliminere oppgaver som bidrar negativt til testens reliabilitet, diskriminerer svakt eller har uønsket vanskelighetsgrad, slik det har blitt foreslått. I tillegg er bakgrunnsvariablene i den første delen av kartleggingsverktøyet kun relevante fra et forskningsøyemed, og kan ekskluderes før det anvendes av naturfagslærere.

Cronbach og Shavelson (2004) mener at for mange anvender Cronbach alfa blindt som et mål på reliabilitet, og anbefaler å legge vekt på standardfeil som et reliabilitetsmål. Standardfeil gir verdier som er enklere å tolke for allmennheten (Cronbach & Shavelson, 2004), og kan gi mer intuitivt mening for lærerne som skal anvende testen. Ferdighetstestens standardfeil er under ett poeng, og det betyr at elevenes sanne score ligger innenfor intervall av to poeng av observert score, gitt at testen er en valid operasjonisering av ferdighetene som undersøkes. Det betyr at fra et reliabilitetsperspektiv er testen egnet til å undersøke enkeltelevers kritiske tenkning. Også sensorreliabiliteten i testen må anses for å være høy, med tanke på at 50 av 53 oppgaver i testen var selvrettende lukkede oppgaver. Det ble i tillegg anvendt konkrete kriterier for poenguttelling i de tre åpne oppgavene.

Dette kan tyde på at testen gir reliable resultater om elevene i utvalg 4 sine sammenlagte ferdigheter til å tenke kritisk, med hensyn til intern konsistens, standardfeil, sensorreliabilitet og korrelasjoner mellom testkategorier og delferdigheter. Høy reliabilitet er en forutsetning for høy validitet (Patel et al., 1995), men det betyr ikke at testen faktisk måler om elevene innehar de enkelte ferdighetene som utgjør elevenes evne til å tenke kritisk. Facione (1990a) identifiserte seks ferdigheter og seksten delferdigheter som er nødvendige for å kunne tenke kritisk. I ferdighetstesten er ferdighetene delt inn i tre testkategorier som består av til sammen åtte delferdigheter. Bakgrunnen for dette valget var å kunne øke antall oppgaver i hver kategori i testen, og dermed få mer reliable data om den kompetansen hver testkategori måler.

Som vist i kapittel 4.4.2 er det utfordringer med reliabiliteten på kategori og delferdighetsnivå. På kategorinivå har kun «Strategier for kritisk tenkning» ( $\alpha=0,78$ ) og «Å trekke velbegrunnede konklusjoner» ( $\alpha=0,69$ ) en intern konsistens over akseptabelt nivå ( $\alpha>0,65$ ), mens verdien på Cronbach alfa for «Forståelse for naturfaglig informasjon» er 0,62. Kun delferdighetene S2 ( $\alpha=0,75$ ) og K3 ( $\alpha=0,66$ ) har akseptabelt nivå av intern konsistens på delferdighetsnivå. Det betyr at fra et reliabilitetsperspektiv er det begrenset hvor mye testen kan fortelle om elevenes ferdigheter på et delferdighetsnivå, fordi scoren i delferdighetene kan

være unøyaktig. For eksempel var det lite intern konsistens i delferdigheten K1 ( $\alpha=0,24$ ), som betyr at testen gir et unøyaktig svar på hvor god evne elever har til å «undersøke og vurdere bevis» gjennom oppgavene tilknyttet delferdigheten. Kategoriene er sammensatt av delferdigheter, og gir et grovere bilde av elevens ferdighetsnivå. Dersom endringene som foreslåes i kategorien «Forståelse for naturfaglig informasjon» implementeres og den interne konsistens i kategorien øker, vil det være akseptabel intern konsistens i alle testkategoriene. I så fall kan score i testkategoriene anvendes for å undersøke hva elevene fikk til og ikke i testen, siden kategoriene undersøker ulike ferdigheter. Dette ville gjort testen mer egnet til å identifisere hvilke ferdigheter enkelte elever mestrer og synes er utfordrende, sammenlignet med testscore.

Sammenlignet med andre tester om kritisk tenkning er ikke resultatene på kategori- og delferdighetsnivå overaskende. Cronbach alfa for de enkelte ferdighetene i California Critical thinking test har blitt målt til å være mellom 0,21 og 0,51 (Leppa, 1997), og Cronbach alfa for ferdighetene i Watson-Glaser test (Form S) ligger mellom 0,17 og 0,74 (Loo & Thorpe, 1999). Testen fra PROBE-prosjektet er innholdsmessig mest lik ferdighetstesten, og Cronbach alfa for de enkelte delferdighetene er målt er mellom 0,30 til 0,74 (Kluge et al., 2018). Denne sammenligningen viser at det er utfordrende å gi reliable mål på ferdighetsnivå innen kritisk tenkning.

Vanskelighetsgraden avgjør hvilket ferdighetsnivå som undersøkes. Testens vanskelighetsgrad ( $p=0,69$ ) er lavere enn optimalt, hvis målet er å differensiere utvalgets kompetanse så godt som mulig (McAlpine, 2002). Det betyr at testen bør gi et mer nyansert bilde av ferdighetene til elevene på lavt ferdighetsnivå, enn elevene på høyt ferdighetsnivå. Vanskelighetsgraden er påvirket at det er færre vanskelige oppgaver, som kan skille mellom respondenter på høyt ferdighetsnivå, enn oppgaver med lav vanskelighetsgrad. Dette er i teorien akseptabelt for en kriteriebasert test (Cohen et al., 2007), hvis elevene innehar de ferdighetene som undersøkes, og oppnår «kriteriene» for å få en høy score. Altså kan dette bety at mange av elevene var kompetente i ferdighetene som testen måler, og derfor var vanskelighetsgraden lav. Dette kan en av årsakene til at resultatene fra ferdighetstesten ikke var normalfordelt i studie 4. Utvalgets størrelse er i tillegg en feilkilde når normalfordeling undersøkes, ettersom sjansen for å oppnå en tilnærmet normalfordeling øker med utvalgets størrelse (Nardi, 2018). Også Kluge et al. (2018) rapporterer om at fordelingen skjevhet mot

høyre i fordelingen av resultater i PROBE-prosjektet, hvor de utviklet en test med lignende fokus som ferdighetstesten.

Det er likevel ikke nødvendigvis tilfelle at ferdighetene som måles i testen er enkle. For eksempel er det tenkelig at oppgavene i del 1 kunne vært vanskeligere hvis jeg hadde valgt andre tekster som elevene skulle vurdere, eller at oppgavene i del 2 ville vært enklere hvis jeg hadde valgt en mindre komplisert case som elevene skulle finne forklaring på. Testen jeg har utformet er svært avhengig av innholdet, og det er utfordrende å finne gode autentiske tekster og problemstillinger som passer for målgruppen. Det faglige innholdet i testen har vært relativt likt gjennom testutviklingen, og det er derfor interessant å se at den gjennomsnittlige vanskelighetsgraden i testen har vært tilnærmet lik i studie 1 ( $p=0,67$ ), studie 3 ( $p=0,66$ ) og studie 4 ( $p=0,69$ ), til tross for at testen har gjennomgått endringer. Dette kan tyde på at valg av naturfaglig innhold i testen påvirker vanskelighetsgraden. Tiruneh et al. (2017) utviklet en test som målte ferdigheter innen kritisk tenkning i fysikk på universitetsnivå, med fokus på elektrisitet og magnetisme, og i studien var testen vanskeligere enn optimalt ( $p<0,5$ ). Elektrisitet og magnetisme kan virke som abstrakte fenomener, og faginnholdet kan altså ha bidratt til den høye vanskelighetsgraden. På en annen side er ikke testen til Tiruneh et al. (2017) direkte sammenlignbar med ferdighetstesten, ettersom testen ikke fokuserte på autentiske tekster og problemstillinger, og det var en eldre målgruppe med bedre forutsetninger for å forstå naturfaglig informasjon.

Det er ikke kun faginnholdet som kan ha påvirket vanskelighetsgraden i ferdighetstesten i studie 4. Den består også av mange AC-oppgaver med kun en distraktor, og slike oppgaver er ofte enklere enn lukkede oppgaver med flere distraktorer eller åpne oppgaver (Haladyna et al., 2002). Bruken av AC-oppgaver kan også være med på å forklare hvorfor det er lite intern konsistens i flere av delferdighetene (Wibowo, 2016). Mange av oppgavene i testen har svært lav vanskelighetsgrad ( $p>0,85$ ). Noen er aktuelle å ekskludere, mens andre anbefaler jeg å beholde, til tross for lav vanskelighetsgrad, fordi de er viktige for oppgavesettet.

En utfordring med testdesignet er at det ikke er mulig å ekskludere alle oppgaver som ga uønskede resultater i item-analysen i studie 4. Årsaken er at testen i stor grad er bygget opp gjennom to oppgavesett med CDIS-oppgaver, altså oppgaver som henger sammen med en stimulus eller annet innhold i testen. Fordi ferdighetstesten skal undersøke hvordan ungdomsskoleelever tenker kritisk rundt autentiske tekster og problemstillinger er de fleste

oppgavene avhengig av en stimulus, og påvirket av elevenes svar på andre oppgaver som er knyttet til samme stimulus. For eksempel kan man anta at elevenes svar på vurderingsspørsmålene i del 1 om kilde 2 er påvirket av deres refleksjoner i oppgavene som handlet om kilde 1. I del 2 er oppgavene sammenknyttet med informasjon om casen, og svar på tidligere oppgaver. Fordelen med å bruke CDIS-oppgaver er at de kan avdekke hvordan elevene tenker gjennom oppgavesettet, kan undersøke komplekse ferdigheter og gjør det mulig å undersøke kompliserte problemstillinger, fordi elevene kan bruke mer tid per stimulus. CDIS-oppgaver er derfor egnet for å undersøke elevens tenkning rundt autentiske naturfagsrelaterte problemstillinger og tekster. For å utvikle slike oppgaver mener jeg at en suksessiv testutvikling er best egnet, slik som det har blitt gjort i dette studiet. På denne måten kan man gjøre endringer på oppgavesettene og se hvordan dette påvirker resultatene. Det er flere ulemper med slike oppgaver. For det første kan elevene misforstå noe tidlig i oppgavesettet, og få følgefeil som påvirker resultatene på videre oppgaver. Dette kan være en reliabilitetsutfordring. I tillegg må skjønnsmessige vurderinger spille en viktig rolle når testen skal revideres med hensyn til item-analysene. Dersom hver oppgave var tematisk isolert kunne jeg ekskludert alle oppgaver som ikke ga ønskelig resultat i item-analyse, slik som det er vanlig å gjøre etter tradisjonelle pilottester, med mange isolerte oppgaver. Testdesignet med CDIS-oppgaver gjør det for eksempel utfordrende å balansere vanskelighetsgraden i testen, ved å fjerne oppgaver med lav vanskelighetsgrad. I tillegg var det kun 22 av 53 oppgaver som diskriminerte på et akseptabelt nivå ( $d > 0,3$ ) i forhold til delferdigheten de skulle måle. Dette påvirket den interne konsistensen i delferdighetene. Spesielt i delferdighetene K1 og K2, som ble foreslått å ekskludere fra testen i kapittel 4.4.2, var det få oppgaver som diskriminerte på et akseptabelt nivå. Uten de 17 oppgavene som tilhører K1 og K2 er det fortsatt kun 55% av de resterende oppgavene som diskriminerer på et akseptabelt nivå. I studie 3 diskriminerte 37 av 51 oppgaver, altså 73%, over akseptabelt nivå, inkludert oppgavene fra delferdighetene K1 og K2. Utvalget i studie 3 var lite ( $n=11$ ), og dette kan være årsaken til at det var så stor forskjell mellom studie 3 og 4. Til sammenligning var det kun 7 av 20 oppgaver som diskriminerte på et akseptabelt nivå i pilottesting som Tiruneh et al. (2017) gjennomførte på testen om kritisk tenkning om magnetisme og elektrisitet. Dette er en utfordring som de anerkjente testene også har, ettersom ingen av testene kan gi mål på ferdighetsnivå med akseptabel intern konsistens (Kluge et al., 2018; O'Hare, 2005).

Med hensyn til intern konsistens og standardfeil vil jeg konkludere med at score på ferdighetstesten som helhet ga et reliabelt mål på elevene i utvalgets ferdigheter til å tenke

kritisk. Reliabilitetsanalysene fra studie 4 indikerer at testen kan være på nivå med eksisterende tester på området, men i denne sammenligningen er ikke utvalgsstørrelse og utvalgstype tatt hensyn til. Det er likevel forbedringspotensiale for testen med hensyn til å balansere vanskelighetsgraden og for å få reliable mål på kategorinivå og delferdighetsnivå i testen. De forslagene til endringer som foreslåes i kapittel 4.4.2 kan bidra til å øke reliabiliteten på de nivåene.

### 5.1.2 Validitet

En reliabel test er verdiløs, om den ikke gir valide mål på det den skal måle (Cohen et al., 2007). Det har ikke blitt gjennomført statistiske tester for å undersøke validiteten til ferdighetstesten. Det betyr at validiteten til testen ikke tallfestes i denne masteroppgaven, og det er behov for mer systematiske analyser for å få et bedre overblikk over testens validitet etter dette masterprosjektet. Jeg vil likevel gjøre rede for hvordan jeg har forsøkt å ta hensyn til begrepsvaliditeten i den versjonen av testen som ble anvendt i studie 4. Dette er gjort gjennom en skjønnsmessig vurdering av samsvaret mellom de teoretiske begrepene og selve måleinstrumentet. Testen er et forsøkt på å operasjonalisere en rekke teoretiske begreper, som ble gjort rede for i kapittel 2.1. Det viktigste teoretiske begrepet for testen er de kognitive ferdighetene som er nødvendige for å tenke kritisk. Under gjøres det rede for hvordan begrepene har blitt forsøkt operasjonalisert og ivaretatt i den endelige ferdighetstesten.

Kritisk tenkning er et viktig begrep i mitt masterprosjekt. Som vist i kapittel 2.1 er det vanlig å knytte kritisk tenkning til evnen til å kunne ta gode beslutninger. Ennis (1985) definerte kritisk tenkning som reflektert og fornuftbasert tenkning som fokuserer på avgjørelser om hva man skal tro eller gjøre. Ferdighetstesten undersøker om elevene kan ta avgjørelser basert på begrenset, men tilstrekkelig, informasjon. I testen skal elevene trekke konklusjoner om to kilders troverdighet, hvor stor kreftisiko som forbindes med spising av kjøtt sammenlignet med røyking og hvilken av seks ferdigformulerte konklusjoner som er den mest sannsynlige forklaringen på en case om et økosystem.

Som vist gjennom masteroppgaven kan kritisk tenkning anses som bestående av ferdigheter og holdninger, og at de anerkjente testene innen kritisk tenkning forholder seg til en slik inndeling. Den konkrete operasjonaliseringen av de kognitive ferdighetene som er nødvendig for å tenke kritisk er i stor grad basert på Facione (1990a). Hensikten med delphistudien som Facione (1990a) gjennomførte var å formulere kjennetegn på den kritiske tenkningen som en



gjennomsnittlig collestudent bør klare, og i studien deles begrepet inn i holdninger og kognitive ferdigheter, som er nødvendige for slik tenkning. Ferdighetstesten skal måle ferdighetene som er nødvendige for å kunne tenke kritisk i naturfag på ungdomsskolenivå. Gjennom de foregående kapitlene har det blitt redegjort for hvordan de seks ferdighetene, og seksten delferdighetene som Facione (1990a) formulerte, har blitt tilpasset faginnholdet, den yngre målgruppen og testformatet. Tilpasningen har ført til at testen måler tre testkategorier og åtte tilknyttede delferdigheter. Jeg har forsøkt å ivareta begrepsvaliditeten gjennom vise åpenhet om hvordan begrepet har blitt operasjonalisert og tilpasset målgruppen og faginnholdet fra ferdighetene slik de ble beskrevet i Facione (1990a), til kategoriene, delferdighetene og de tilknyttede oppgavene som anvendes for testen. Ifølge Ku (2009) er det viktig å variere oppgavetyper når man skal måle et så sammensatt begrep som kritisk tenkning. I ferdighetstesten er det derfor anvendt ulike former for lukkede oppgaver og noen åpne oppgaver for å ivareta innholdsvaliditeten knyttet til begrepet.

Et spørsmål som bør stilles er om deler av begrepet «kognitive ferdigheter tilknyttet kritisk tenkning» slik Facione (1990a) beskrev ferdighetene har blitt tapt i operasjonaliseringen. Som vist til i kapittel 3 måtte ferdighetene tilpasses testens omfang og målgruppens yngre alder. Jeg har forsøkt å ivareta ferdighetene gjennom testutviklingen, og elevene må gjennom oppgavene *tolke, analysere, vurdere, trekke konklusjoner og forklare* for å få poeng i testen. Den siste ferdigheten som Facione (1990a) anvender, «*selvregulering*» ønsker jeg å argumentere for er kun delvis målt, ettersom elevene må reflektere over hvordan de tenkte for å komme frem til en konklusjon og anvende strategier for å komme frem til riktig svar, men testen måler ikke direkte om elevene kan vurdere kritisk hva som kan påvirke egne meninger eller korrigere mangler ved egne resonnementer.

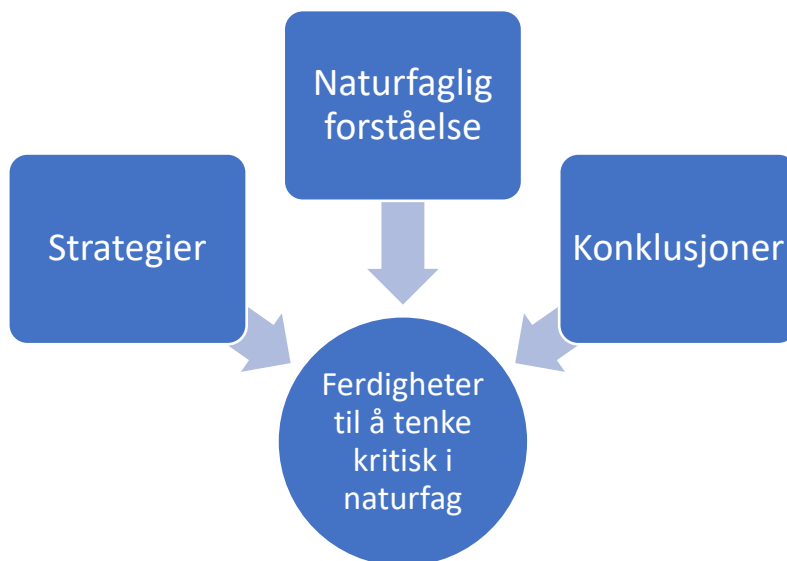
Som gjort rede for i kapittel 2.1, er det ingen selvfølge at det er de samme ferdighetene som er relevante for å tenke kritisk innen alle fagdisipliner. Vieira og Tenreiro-Vieira (2016) mener at det å tenke kritisk i naturfag handler om å kunne kombinere de generelle ferdighetene og holdningene med naturfaglige kunnskaper og naturvitenskapelige kriterier. Testen tar derfor utgangspunkt i naturvitenskapelige temaer, og selv om elevene blir introdusert for nødvendig naturfaglig kunnskap gjennom testen, er de avhengig av å kunne forstå og anvende den. Elevene må for eksempel i del 2 av testen både forstå den informasjonen som blir gitt om økosystemer og kunne anvende dette for å logisk resonere seg frem til hvordan endringer i økosystemet utenfor Alaska påvirker artene som lever der. Vieira og Tenreiro-Vieira (2016)

kombinerer både Facione (1990a), som presenterer generelle kjennetegn på kritisk tenkning og Bailin (2002), som mener kritisk tenkning må være knyttet til temaspesifikke kunnskaper. Ved å anvende en ferdighetstest som er knyttet opp mot naturfaglige temaer og problemstillinger, unngår ferdighetstesten noen av de punktene Bailin (2002) mener er feiloppfatninger av kritisk tenkning.

Å kunne tolke og vurdere argumenter er en viktig egenskap innen naturvitenskapen, og sentralt ved kritisk tenkning. Å vurdere argumenter var derfor ved studie 1 og 2 en egen testkategori og en egen del av testen med 14 oppgaver, men ble redusert til fem oppgaver tilknyttet delferdigheten K4. Dette ble gjort for å begrense omfanget av testen, og fordi oppgavene var problematiske gjennom studie 1 og 2, selv etter at de ble forsøkt revidert. Å kunne vurdere argumenter er en del av de ferdighetene som ble identifisert som viktige for å kunne tenke kritisk ifølge Facione (1990a). Det å kunne vurdere et argument handler i hovedsak om å kunne vurdere om argumentet er logisk gyldig og hvor relevant argumentet er til situasjonen det anvendes i. Elevene må i oppgavene vurdere relevansen til fire argumenter i del 1.3, men må kun vurdere et arguments logiske konsekvenser i en oppgave (1.1 j). Testen er derfor ikke egnet til å gi en score på elevenes ferdighet til å vurdere argumenter. Denne ferdigheten kan likevel ses i sammenheng med kategorien «å kunne konkludere», som gir en score på hvor egnede elevene er til å tolke og vurdere bevis og premisser for å trekke en logisk og velbegrunnet konklusjon.

En viktig del av det å kunne tenke kritisk i dagens informasjonssamfunn handler om å kunne lese kritisk. Facione (1990a) inkluderer det å kunne vurdere troverdigheten til kilder og påstander som en delferdighet innen kritisk tenkning, og Ferguson og Bråten (2013) anser kildekritikk som et av tre mest sentrale elementene av kritisk tenkning. Derfor har kildekritikk fått en stor rolle i ferdighetstesten. Ifølge Weyergang og Frønes (2020) kan elever anses som kritiske lesere hvis de kan stille seg utenfor en tekst, og stille enkle spørsmål om teksten. Gjennom del 1 i testen må elevene redegjøre for hvilke strategier de anvender for å undersøke troverdigheten til en tekst på internett, og vurdere to teksters troverdighet. Elevenes kritiske lesning blir altså undersøkt gjennom testen. Utfordringen er at testen kun måler elevenes vurdering av to tekster, og valget av tekster blir derfor viktig for elevenes resultater, og oppgavenes vanskelighetsgrad.

Oppsummert er innholdet og oppgavene i testen knyttet til de teoretiske begrepene. Testen skal kunne produsere en score som representerer elevenes samlede ferdigheter til å tenke kritisk i naturfag. Som vist over indikerer kvalitetsmarkørene at testen gir et reliabelt mål i utvalg 4. Testscoren er en sammenlagt score på alle oppgavene tilknyttet delferdighetene og kategoriene i testen. Jeg vil trekke frem to argumenter for at testscore kan anvendes som et validt mål på elevenes ferdigheter til å tenke kritisk i naturfag. For det første har testen i hovedsak tatt utgangspunkt i Facione (1990a) i utformingen av kategoriene og delferdighetene i testen, som knytter ferdighetene opp mot begrepet kritisk tenkning. Det betyr at summen av ferdighetene vil konstituere en slags kognitiv prosesseringskraft som skal til for å tenke kritisk. I min test er ferdighetene sammenfattet i tre kategorier, og sammenlagt score på de tre kategoriene er den sammenfattede testscoren, om vist i Figur 5-1. For det andre er det tydelig samsvar mellom testscore og delferdighetene som måles i testen, med unntak av K4, og enda tydeligere korrelasjoner mellom testscore og kategoriene i testen. I tillegg setter reliabiliteten en grense for validiteten, og reliabiliteten til ferdighetstesten for utvalg 4 virker til å være sammenlignbart med tilsvarende tester på kritisk tenkning.



Figur 5-1: Testscore er sammenfattet fra score i de tre kategoriene.

Hvordan poengene er fordelt på testelementene er et aspekt ved innholdsvaliditet. Hvis alle ferdighetene anses som like viktige, bør de også vektlegges like mye i poengutdelingen. Dette er det ikke tatt hensyn for i testutviklingen, og enkelte ferdigheter og delferdigheter er knyttet til flere oppgaver enn andre. Hovedfokuset i testen er de ferdighetene som er relevante for å kunne trekke velbegrunnede konklusjoner og å kunne være kildekritisk, men alle ferdighetene er representert gjennom de tre kategoriene, i ulik grad. Det største problemet med vektingen i

ferdighetstesten er antallet poeng knyttet til kategoriene. Maksimal score i kategorien «Strategier for kritisk tenkning» er 19 poeng. Testens minste kategori i forhold til oppgaver og poengvekting er «Naturfaglig forståelse» med 12 poeng. Testkategorien «Å trekke velgrunnede konklusjoner» er den mest omfattende med 27 poeng. I kapittel 4.4.2 foreslås det å øke omfanget av kategorien «Naturfaglig forståelse», og å redusere antallet oppgaver i kategorien «Å trekke velgrunnede konklusjoner». Denne endringen vil kunne virke positivt for innholdsvaliditeten, og testscore vil bli et bedre mål på elevenes ferdigheter til å tenke kritisk i naturfag.

Validitet kan estimeres gjennom faktoranalyse eller gjennom å sammenligne egen test opp mot en kjent test som man vet har høy validitet. Faktoranalyse har ikke blitt gjennomført i dette masterprosjektet, av hensyn til prosjektets omfang. Verken utviklerne i California Critical thinking test eller Watson-Glaser test oppgir mål på faktoranalyse, men uavhengige studier viser til validitetsutfordringer i målene fra faktoranalyse for de enkelte ferdighetene (Kluge et al., 2018). Ifølge O'Hare (2005) er begrepsvaliditeten ivaretatt i testene gjennom å utvikle testene med utgangspunkt i den teoretiske diskusjonen om kritisk tenkning, slik jeg har forsøkt å gjøre. Heller ikke resultatene fra faktoranalyse viser at man kan skille mellom de ulike ferdighetene i PROBE-testen, og Kluge et al. (2018) mener årsaken til dette er at ferdighetene ikke var endimensjonale og adskilte, i tillegg til at det var høyere intern konsistens mellom de ulike tematiske delene av testen ( $\alpha=0,68-0,77$ ), enn mellom ferdighetene ( $\alpha=0,30-0,74$ ). Da jeg undersøkte den interne konsistensen i ferdighetstesten, viste analyse av Cronbach alfa at del 1 ( $\alpha=0,83$ ) og del 2 ( $\alpha=0,73$ ) var mer konsistente enn testens kategorier ( $\alpha=0,62-0,78$ ). Årsaken kan være at det er flere oppgaver knyttet til de to delene enn testkategoriene, eller at ferdighetene som måles i testen kan være mer sammenflettet enn forventet, slik Kluge et al. (2018) foreslår. Kluge et al. (2018) mener derfor at tester innen kritisk tenkning er best målt som en testscore, enn på ferdighetsnivå. Denne konklusjonen mener jeg er relevant for ferdighetstesten også, med utgangspunkt i drøftingen av testens reliabilitet og validitet. Likevel er det mitt håp at de foreslåtte revideringene av testen kan føre til mer intern konsistens i kategoriene. Det kan bety at score på kategorinivå kan gi et grovt bilde av hvilke ferdigheter eleven er kompetent i, og hvilke som bør styrkes.

Angående ferdighetstestens reliabilitet er mitt hovedpoeng at den virker til å være på høyde med de eksisterende testene innen kritisk tenkning, sett fra et psykometrisk standpunkt. Som vist over kan heller ingen av de nevnte studiene på testenes kvalitet vise til positive

validitetsmål. Men som O'Hare (2005) konkluderte i sin gjennomgang av de mest anerkjente testene på området, imponerer ikke de eksisterende testene med sin psykometriske kvalitet. Likevel bør validiteten til ferdighetstesten undersøkes med mer systematiske analyser for å få bedre innsikt i hvor treffsikkert mål testen er på elevenes ferdigheter til å tenke kritisk. Det har ikke blitt tatt hensyn til utvalgsstørrelse eller rekruteringsprosedyre i sammenligningen med andre tester. I utprøvingen av PROBE-testen undersøkte for eksempel Kluge et al. (2018) et utvalg som var over tre ganger større enn utvalg 4, som var rekruttert fra flere skoler. Det er derfor behov for å utprøve ferdighetstesten på et større tilfeldig utvalg for å kunne sammenligne resultatene med andre tester med større sikkerhet.

Over har jeg gjennomgått både styrkene ved kvaliteten på ferdighetstesten, og utfordringer som bør adresseres for at testen skal gi mer reliable og valide resultater. Målet bør være å videre redusere feilkildene ved den observerte scoren, slik at det i større grad er elevenes sanne ferdigheter som observeres. Slike standardiserte tester vil likevel aldri gi et komplett bilde av en elevs tenkning, men bør anvendes for å komplementere den kvalitative oppfattelsen lærere har av elevenes kompetanse.

## 5.2 Testens resultater

Dette masterprosjektet har to forskningsspørsmål. De er knyttet til hverandre, og til oppgavens hensikt. Forskningsspørsmål 2 spør hva som kjennetegner testens resultater, og dette er avhengig av testens kvalitet, og kvaliteten på variablene i det resterende kartleggingsverktøyet. Formulert annerledes blir spørsmålet: «Hvilken informasjon om utvalg 4 kan resultatene på kartleggingsverktøyet gi?».

I kapittel 4.4.3 presenteres vanskelighetsgraden til de ulike delferdighetene i testen, og som nevnt over er vanskelighetsgraden til delferdighetene, slik de måles i testen, påvirket av innholdet, oppgaveformulering og oppgavetype. Det betyr at jeg ikke kan påstå at det er enklere å vurdere troverdigheten til en kilde, som S2 måler, enn å kunne trekke velbegrunnede konklusjoner (K3). Alt det er mulig å si med sikkerhet er at flere i utvalget fikk høy score i delferdigheten S2 ( $p=0,84$ ) enn i K3 ( $p=0,69$ ) i ferdighetstesten, med de oppgavene og det innholdet som ble anvendt. Delferdigheten S2 er et eksempel på hvordan de valgene jeg tok i testdesignet påvirket testens resultater. At S2 utmerker seg som spesielt overkommelig for elevene i testen, bør for eksempel ses i sammenheng med at oppgavene som ble anvendt for å måle delferdigheten i stor grad var AC-oppgaver, som ofte har lavere vanskelighetsgrad. I

tillegg blir elevene implisitt presentert for kriterier for hva som kjennetegner en troverdig kilde gjennom oppgaveformuleringene. Det er derfor sannsynlig at vanskelighetsgraden kunne vært høyere om det ble anvendt åpne oppgaver for å måle delferdigheten. Resultatet kunne også vært annerledes hvis testen hadde tatt utgangspunkt i andre tekster, med mer komplekse problemstillinger om troverdighet.

Gjennom PROBE-prosjektet fant også Kluge et al. (2018) at de oppgavene som undersøkte elevenes evne til å forstå den naturfaglige informasjonen i testen hadde lavest vanskelighetsgrad, mens de oppgavene som undersøkte elevenes evne til å begrunne valgene sine var testens vanskeligste. Som vist i kapittel 4.4.3 stemmer dette overens med resultatene på ferdighetstesten. Begge delferdighetene som måler elevenes forståelse av det naturfaglige innholdet ( $p=0,73-0,75$ ) var blant de tre delferdighetene med lavest vanskelighetsgrad. S1, som måler elevenes evne til å begrunne og forklare hvilke strategier de anvender, var delferdigheten som har nest høyest vanskelighetsgrad ( $p=0,59$ ). Spørsmålet blir om det er en reel forskjell i hvor vanskelige delferdighetene er for elevene i utvalget, eller kan det være slik at det var mer utfordrende for elevene å svare på åpne oppgaver, som S1 var sammensatt av? Dersom det er sistnevnte forklaring som er riktig kan det være at de åpne oppgavene har en høyere vanskelighetsgrad fordi de gir en bredere innsikt i hva respondenten kan, slik Messick (1993) argumenterer for. Om det er slik at enkelte av delferdighetene i testen i virkeligheten er mer utfordrende enn det testens resultater indikerer, bør de lukkede oppgavene revideres, ettersom Haladyna et al. (2002) argumenterer for at hva som kan måles med lukkede oppgaver kun avhenger av hvor kompetente de som utvikler testen er.

Med utgangspunkt i testens resultater har jeg utformet kjennetegn for hva som var utfordrende og enkelt for elever på de ulike ferdighetsnivåene i testen, som vises i Tabell 5-1. I kapittel 5.1.1 ble det gjort rede for at resultatene i flere av delferdighetene ikke er konsistente nok til å gi reliable resultater. Dette er en feilkilde ved kjennetegnene slik de er formulert i Tabell 5-1. Kjennetegnene har tatt utgangspunkt i delferdighetenes vanskelighetsgrad, og korrelasjonen mellom delferdighetene og testscore.

Tabell 5-1: Kjennetegn på de ulike ferdighetsnivåene.

Kjennetegn	Høyt ferdighetsnivå	Middels ferdighetsnivå	Lavt ferdighetsnivå
Vurdere troverdigheten til kilder ut fra gitte kriterier (S2)	X	X	X
Avkode og forstå naturvitenskaplig informasjon i ulike former (N1 + N2)	X	X	
Kunne trekke velbegrunnede og logiske konklusjoner, gjennom å utelukke andre mulige konklusjoner basert på informasjon (K1, K2 og K3)	X		
Gjøre rede for og formulere egne strategier for å trekke gode konklusjoner (S1)	X		
Vurdere hvor relevante argumenter er for en problemstilling (K4)			

Kjennetegnene over er generelle og ikke representative for alle elevene i utvalg 4, ettersom ferdighetsnivåene tok utgangspunkt i testscore, og ingen av delferdighetene korrelerte perfekt med testscore. For eksempel var det enkelte elever på lavt ferdighetsnivå som fikk til å «*avkode og forstå naturfaglig informasjon i ulike former*», men vanskelighetsgraden på N1 og N2, og den tydelige korrelasjonen de to delferdighetene har til testscore, indikerer at elevene på høyt og middels ferdighetsnivå fikk dette til i større grad enn de på lavt ferdighetsnivå. Den delferdigheten som korrelerte svakest med testscore var K4, og dette tilsier at forskjellen mellom ferdighetsnivåene ikke er like stor som ved de andre delferdighetene. Årsaken til dette kan være at selve evnen til å vurdere argumenter ikke har noen sammenheng med resten av delferdighetene som utgjør elevenes ferdigheter til å tenke kritisk, eller at det var utfordringer med kvaliteten til oppgavene. Dette var den delferdigheten som elevene på høyt ferdighetsnivå gjennomsnittlig scoret svakest på, og betyr at K4 var utfordrende for elever på høyt ferdighetsnivå. Kjennetegnene i Tabell 5-1 kan anvendes av lærere for å veilede elever på ulike ferdighetsnivåer, med utgangspunkt i deres testscore.

Resultatene fra testen viste ikke signifikant korrelasjon ( $r=0,13$ ,  $p=0,29$ ) mellom gjennomføringstid og ferdighetsscore. I alle tre ferdighetsnivåene var det stor variasjon i hvor lang mye tid elevene brukte på kartleggingsverktøyet. Mange faktorer kan ha påvirket dette. Enkelte elever på høyt ferdighetsnivå kan ha vært raske lesere, og gjennomførte testen raskt, mens andre kan ha fått høy score gjennom å lese oppgavene nøye. Elevene på lavt ferdighetsnivå kan både bestå av elever som mistet motivasjonen tidlig i testen, og svarte uten å reflektere over oppgavene, og dermed gjennomførte testen raskt, i tillegg til elever som var svake lesere, og dermed brukte lang tid på testen. Resultatene stemmer overens med

resultatene på PROBE-prosjektet, hvor Kluge et al. (2018) heller ikke fant sammenheng mellom gjennomføringstid og score i testen.

Kartleggingsverktøyet undersøker både bakgrunnsvariabler og elevenes holdninger, i tillegg til de ferdighetene som blir målt i testen. Bakgrunnsvariablene ble anvendt for å undersøke om testen kan skille ulike elevgrupper. Resultatene i kapittel 4.4.3 viser at det var flere variabler som ga signifikante forskjeller i gjennomsnittet mellom grupper i utvalget. Ettersom utvalget ikke er tilfeldig utvalgt kan det ikke tas utgangspunkt i at resultatene kan generaliseres for populasjonen. Likevel kan signifikante forskjeller i gjennomsnitt indikere at det er mer sannsynlig at det faktisk er en forskjell i befolkningen. Elevene i utvalget som hadde minst en forelder med nordisk opprinnelse scoret i gjennomsnitt 8,04 poeng mer enn de som ikke har nordiske foreldre. For utvalget var denne forskjellen signifikant ( $p < 0,01$ ). Resultatet kan likevel ses i sammenheng med Weyergang og Frønes (2020) sin analyse av elevers evne til kritisk lesning. Der poengterer de at skolen ikke er gode nok på å systematisk lære elevene om hvordan de skal lese tekster kritisk, og at hjemmets tekstpraksis blir viktig. Hvis elever med minoritetsspråklig bakgrunn vokser opp med andre tekstpraksiser kan dette påvirke scoren i ferdighetstesten, og dette kan være en mulig forklaring på hvorfor det er en forskjell i gjennomsnittet. I tillegg er ordforråd og norskkompetanse viktig for å kunne forstå og navigere en skriftlig test. Ved skolen utvalget er rekruttert fra er det høy andel med minoritetsspråklige elever. Dersom det er slik at minoritetsspråklige elever generelt scorer lavere enn etnisk norske elever kan dette bety at testen vil ha en lavere vanskelighetsgrad hvis den ble utprøvd på en skole utenfor Oslo, med en mer etnisk homogen elevgruppe. Score på ferdighetstesten samsvarer også med den selvrapporterte karakteren i både naturfag og norsk, og det er en signifikant forskjell i gjennomsnittet ( $p < 0,01$ ) mellom elever som fikk 3 eller 4 og elevene som fikk 5 eller 6. Det kan være flere årsaker til dette. Det kan være at elever som generelt får høye karakterer på skolen gjennomsnittlig har bedre evne til å tenke kritisk, eller at kompetansen elevene anvender i norsk og i naturfag er relevante for ferdighetstesten. Det kan også bety at de generelt mestrer å navigere i skriftlige tester bedre enn elever med lave karakterer. Det er mulig at elever på lavt og middels ferdighetsnivå for eksempel mestrer ferdighetene som testen skal undersøke, men svarer feil grunnet at de leser feil, misforstår begreper eller av andre grunner har utfordringer med skriftlige tester.

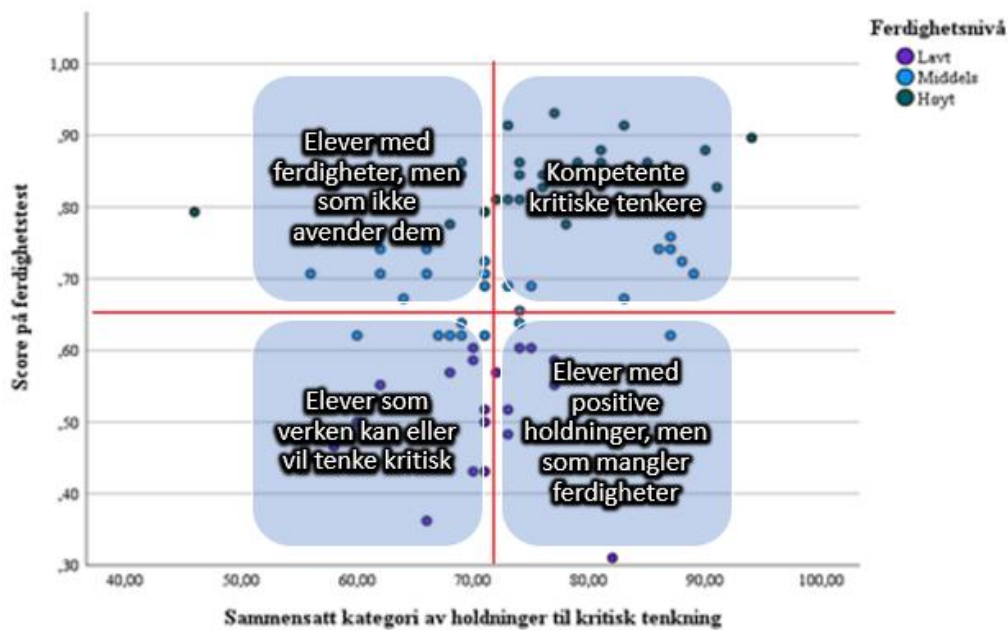
Det mest overraskende resultatet var hvor stor forskjell det var i gjennomsnittet på ferdighetstesten til elevene som svarte «ja» på spørsmålet «*Tror du det er like*



*kreftfremkallende å spise kjøtt som å røyke*», og de som svarte «*nei*». Spørsmålet var en del av spørreskjemaet elevene fylte ut før de startet på testen. De 10 elevene som mente at kjøtt var like kreftfremkallende som røyking scoret i gjennomsnitt 11,27 poeng lavere enn de som ikke trodde det. Dette er den største forskjellen i gjennomsnitt som ble funnet i alle variablene som ble undersøkt, og forskjellen var signifikant. Det er derfor relevant å reflektere over hva som kan være årsaken til denne forskjellen. Er det slik at de elevene som tror at røyk og kjøtt er like kreftfremkallende fikk lavere score i testen, fordi budskapet i kilde 1 samsvarte med deres egne tanker, og derfor svarte feil på oppgavene tilknyttet del 1? Eller kan det være slik at spørsmålet måler noe knyttet til deres evne til å tenke kritisk, som å kunne se logiske konsekvenser av gitte premisser? Forklaringen kan også være at dette var elever som ikke leste spørsmålet nøye eller ikke forsto innholdet i spørsmålet. I så fall er det ikke usannsynlig at dette er de samme elevene som hadde lavere forståelse for oppgavene, ettersom de ikke forsto dem eller fordi de var lite nøye. Hvis dette er tilfellet, ville resultatene vært annerledes om spørsmålet ble stilt muntlig.

Det er en svak, men signifikant korrelasjon ( $r_s=0,33$ ,  $p=0,01$ ) mellom testscore og den sammensatte kategorien «holdninger knyttet til kritisk tenkning». Resultatene viser at holdninger til kritisk tenkning ikke har stor betydning for hvilke ferdigheter elevene i utvalget har til å tenke kritisk, men selv om det er stor variasjon i utvalget har en større andel elever på høyt ferdighetsnivå positive holdninger til kritisk tenkning enn de på lavt ferdighetsnivå. Facione (2015) kaller slike holdninger «*a critical spirit*», hvor holdningene bestemmer om og når ferdighetene til å tenke kritisk blir anvendt. Elever som har svært positive holdninger til kritisk tenkning, vil altså bruke ferdighetene sine til å tenke kritisk i flere situasjoner. For å tenke kritisk må man stoppe opp, granske den tilgjengelige informasjonen og reflektere over hvilken avgjørelse som passer best til situasjonen. Ifølge Facione (2015) foregår slik tenking i det «trege kognitive systemet». Det betyr at elevens holdninger til kritisk tenkning avgjør når elevene skal anvende det raske eller trege kognitive systemet. Fra et slikt perspektiv vil elevene som ligger øverst til høyre i Figur 5-2, altså elever som både scoret høyt på ferdighetstesten og har positive holdninger til kritisk tenkning, være kapable og gjenkjenne situasjoner å anvende det trege systemet til å kritisk analysere situasjonen. De er altså kompetente kritiske tenkere. Dette er i teorien ikke tilfellet for de tre andre kategoriene i Figur 5-2, ettersom de enten vil unngå å bruke det trege systemet, eller at de har begrensede ferdigheter til å tenke kritisk. Tabell 5-2 gir en oversikt over de teoretiske kjennetegnene til

elevene med utgangspunkt i score på ferdighetstesten og holdningene til kritisk tenkning.



Figur 5-2: En teoretisk oversikt hvilke elever som innehar kompetansen til å tenke kritisk ut fra forståelsen om at kritisk tenkning består av holdninger og ferdigheter. Oversikten har tatt utgangspunkt i figur 4.12.

	Negative holdninger til kritisk tenkning	Positive holdninger til kritisk tenkning
Høy score i ferdighetstesten	<ul style="list-style-type: none"> <li>- Kapabel til å tenke kritisk i naturfag</li> <li>- Vil anvende det raske systemet hvis det er mulig</li> </ul>	<ul style="list-style-type: none"> <li>- Kapabel til å tenke kritisk i naturfag</li> <li>- Vil anvende det trege systemet når det er mulig.</li> </ul>
Lav score i ferdighetstesten	<ul style="list-style-type: none"> <li>- Begrensede ferdigheter til å tenke kritisk</li> <li>- Vil anvende det raske systemet hvis det er mulig</li> </ul>	<ul style="list-style-type: none"> <li>- Begrensede ferdigheter til å tenke kritisk</li> <li>- Vil anvende det trege systemet når det er mulig.</li> </ul>

Tabell 5-2: Kjennetegn på ulike elevgruppers kompetanse i kritisk tenkning.

Å generalisere slik om elevenes liv, utenfor det som har blitt observert i kartleggingsverktøyet, bør man være forsiktig med. Premissene for en slik konklusjon er at elevenes virkelige kompetanse til å tenke kritisk er målt gjennom ferdighetstesten og holdningsvariablene. Et viktig poeng i denne masteroppgaven er at det er knyttet usikkerhet til målene. Det er i tillegg gjort mindre kvalitetssikring på holdningsvariablene enn ferdighetstesten, som betyr at det er mer usikkert om holdningsvariablene gir valide og reliable mål på elevenes disposisjon til å tenke kritisk. Det var kun intern konsistens i den sammensatte kategorien om kritisk tenkning, mens de enkelte holdningskategoriene knyttet til kritisk tenkning ikke ga konsistente resultater. Det er også et premiss at det trege kognitive systemet er avhengig av ferdigheter og holdninger, altså at forklaringen til Facione (2015)

stemmer. Ifølge Facione (1990a) finnes det temaspesifikke holdninger, i tillegg til de generelle. Det betyr at elevene kan være generelt lite opptatt av å analysere påstander og situasjoner kritisk, men kan være interessert i å gjøre det innenfor enkelte områder. Figur 5-2 må derfor leses med flere forbehold. I tillegg er det i figuren en tydelig grense mellom de fire gruppene av elever, men dette er ikke tilfelle i virkeligheten. Grensene i Figur 5-2 er ikke absolutte, men ment for å illustrere et mulig skille. Det betyr at respondentene som er nærmest det nederste venstre hjørnet fikk betydelig svakere testscore og svarte mer negativt på holdningsvariablene enn elever som befinner seg nærmere spredningsplottets midtpunkt. Figur 5-2 bør derfor anses som en teoretisk illustrasjon på forholdet mellom ferdigheter og holdninger, og kan ikke anvendes for å forutsi enkeltelevers evne og vilje til å tenke kritisk.

### 5.3 Kritisk blikk på masterprosjektets metode

Hvor mye vekt som bør legges på resultatene fra masterprosjektet er avhengig av metoden og utvalget. Det er derfor viktig å ta et kritisk blikk på metoden som er anvendt i masteroppgaven. Den suksessive testutviklingen som er anvendt gjennom å bruke flere studier gir prosjektet en økt fleksibilitet sammenlignet med en tradisjonell pilotstudie. Grunnen til dette er at det har vært mulig å revidere måleinstrumentet mellom hver studie. Ulempen med å anvende suksessiv testutvikling er at det er en tidskrevende prosess, som produserer mye data som må behandles og redegjøres for i oppgaven. I tillegg er det begrenset hvor omfattende studiene kan være med hensyn til utvalgsstørrelse, ettersom respondentene til hver studie må rekrutteres og undersøkelsen må gjennomføres. Det er ikke anvendt et tilfeldig utvalg, som er nødvendig for å kunne generalisere for populasjonen. Dette begrenser denne studiens mulighet til å komme med generaliserte slutninger om kartleggingsverktøyets kvalitet, og hva resultatene på kartleggingsverktøyet kan fortelle om elevers kritiske tenkning. Derfor må alle resultater som er presentert tolkes med forsiktighet, og kun knyttes til utvalgene som er anvendt i studien. Det er likevel grunn til å tro at utvalget kan være representativt for skoler med lignende elevgrupper, til tross at det ikke er tilfeldig utvalgt. Det er for eksempel ingen faglige vurderinger som lå til grunn for de klassene som ble valgt ut til utvalg 4. Elevgruppene i utvalgene var valgt fra samme skole. Skolen har et stort mangfold blant elevene, og resultatene bør derfor kunne gi innsikt i både hva elever med ulike forutsetninger og på ulike faglige nivåer får til. Det er en styrke at ferdighetstesten, har blitt undersøkt gjennom fire studier, og de observasjoner som er felles på tvers av studiene bør anses som styrket, som for eksempel testens vanskelighetsgrad. I tillegg var utvalg 4 innenfor den

utvalgsstørrelsen Cohen et al. (2007) anbefaler å anvende ved en pilotstudie, til tross for at fraværet var høyere enn normalt i klassene i utvalgene grunnet Covid-19. De resultatene som har blitt drøftet i dette kapitlet kan derfor anses som indikasjoner som bør undersøkes nærmere med videre studier. Frafallet fra utvalg 4 kan være en feilkilde, og det kan være at det ikke var tilfeldig hvilke elever som ikke kunne eller valgte å ikke delta i studien. Dette kan altså ha påvirket resultatene. Jeg var ikke til stede når tre av de fire klassene gjennomførte undersøkelsen i studie 4, grunnet Covid-19-situasjonen, og dette gjør at det kan være feilkilder knyttet til gjennomføringen som jeg ikke kjenner til. Det ble heller ikke tilrettelagt for enkeltelever som tok testen, og dette kan ha bidratt til at feilkilder påvirket den observerte kompetansen til å tenke kritisk som elevene fikk vist.

Som vist i kapittel 2 finnes det mange meninger om hva kritisk tenkning innebærer. Dette studiet har i hovedsak belaget seg på Faciones (1990a) tolkning av begrepet. Det betyr at ferdighetstestens begrepsvaliditet er knyttet til dette perspektivet på kritisk tenkning. Ut fra divergerende syn på kritisk tenkning vil ikke testen være valid. Det har også vært behov for tilpasse de ferdighetene som Facione (1990a) identifiserte, fra collegenivå til ungdomsskolenivå, og i denne tilpasningen kan sentrale elementer av kritisk tenkning gått tapt. Gjennom masteroppgaven har jeg derfor forsøkt å vise hvordan jeg har tenkt når de teoretiske begrepene har blitt operasjonalisert i måleinstrumentet, og mye arbeid har blitt lagt inn i kvalitetssikring av ferdighetstesten.

Gjennom å studiene har jeg forsøkt å få innsikt i elevenes tolkning og forståelse av oppgavene, og gjøre revideringer med hensyn til både reliabilitet og validitet. Dette har gjort at jeg kunne utvikle sammenhengende oppgavesett med CDIS-oppgaver. Dette ville vært utfordrende å få til gjennom et tradisjonelt pilotstudiedesign. Gjennom testutviklingen har jeg forsøkt å følge anbefalinger knyttet til ulike kvalitetsmarkører for testen, som diskrimineringsevne, vanskelighetsgrad, standardfeil og intern konsistens. Samtidig som testdesignet muliggjør at jeg kan belage testen på autentiske tekster og problemstillinger knyttet til naturvitenskapelig informasjon, har det vært mer utfordrende å ekskludere oppgaver basert på item-analyse, sammenlignet med en tradisjonell pilotstudie. Grunnen til dette er at elevenes svar på CDIS-oppgavene i testen er påvirket av andre oppgaver som knyttet til samme stimulus. Dette har ført til at jeg ikke foreslo i kapittel 4.4.2 å ekskludere alle oppgaver med svak diskrimineringsevne eller med svært lav vanskelighetsgrad. Det betyr at etter dette masterprosjektet er det forbedringspotensiale for ferdighetstesten. Et testdesign

basert på isolerte oppgaver ville sannsynligvis unngått denne problematikken, men det ville gått på bekostning av hvor informasjonstunge problemstillinger som kunne undersøkes.

Ettersom ferdighetstesten har fått mest oppmerksomhet gjennom dette masterprosjektet, er det grunn til å være mer kritiske til resultatene fra de to spørreskjemaene i kartleggingsverktøyet. Informasjonen som ble innhentet om elevene i studiet var anonym, og avhengig av selvrapportering. Det er flere utfordringer ved selvrapportering, som for eksempel at respondentene kan mangle selvinnsett eller at hukommelsen svikter (Ringdal, 2018). Det kan også være at de bevisst ikke prøver sitt beste eller svarer useriøst, slik som de besvarelsene som ble ekskludert fra studien etter datarensing. For eksempel var det uventet at elevene i utvalget som svarte at de «forsto alt» sist gang de leste i en lærebok i naturfag fikk gjennomsnittlig lavere testscore enn de svarte at de «forsto det meste». Dette resultatet kan tyde på at variabelen ikke er et godt mål på leseforståelse. Det ville styrket studien om det ble anvendt flere kjente eller utprøvde variabler fra andre studier. Dette er gjeldene ved alle de tre delene av kartleggingsverktøyet, selv om enkelte holdningsvariabler tok utgangspunkt i tilgjengelige eksempelspørsmål.

Å utvikle gode holdningsvariabler er utfordrende, og spørreskjema om holdninger ble kun utprøvd i studie 3 og 4. Flere av holdningskategoriene hadde intern konsistens under det nivået denne studien har satt som akseptabelt. Dersom kartleggingsverktøyet skal videreutvikles, og det er et mål å undersøke helheten av elevenes kompetanse til å tenke kritisk, er det nødvendig å revidere eller erstatte holdningskategoriene.

## 6 Konklusjon

Fremtiden vil trenge voksne som kan tenke kritisk om naturvitenskapelige temaer og problemstillinger. Dette er nødvendig både for å mestre naturfaget, men også fordi kommende generasjoner må kunne ta veloverveide avgjørelser rundt sosiovitenskapelige kontroverser, og manøvrere seg gjennom informasjonssamfunnet. Hvis naturfagslærere skal videreutvikle slik tenkning hos elevene må de ha innsikt i hva elevene mestrer, og hva de synes er utfordrende. Det finnes få tester av kritisk tenkning som undersøker elevenes ferdigheter til å tenke kritisk rundt autentiske tekster og problemstillinger med naturvitenskapelig innhold. Derfor er hensikten med dette masterprosjektet å utvikle og teste et kartleggingsverktøy for å måle om ungdomsskoleelever har de ferdighetene som er viktige for å kunne tenke kritisk i naturfag, med fokus på autentiske naturvitenskapelige problemstillinger. Kartleggingsverktøyet ble operasjonalisert og tilpasset målgruppen med utgangspunkt i Faciones (1990a) konkretisering av kritisk tenkning, og ble videreutviklet gjennom fire suksessive studier, hvor hovedfokuset var å kvalitetssikre ferdighetstesten (FS1).

Den ferdighetstesten som ble anvendt i studie 4 har en reliabilitet som kan virke sammenlignbar med anerkjente og eksisterende tester som undersøker kritisk tenkning, med utgangspunkt i resultatene fra utvalg 4 på de psykometriske kvalitetsmarkørene som er anvendt i dette prosjektet. Testens validitet er ikke undersøkt like systematisk som reliabiliteten, men jeg forsøkt å ivareta begrepsvaliditeten gjennom åpenhet om valgene jeg har gjort i operasjonaliseringen og testutviklingen. Testen kan derfor anses som anvendelig for å få innsikt i elevenes ferdigheter, med visse forbehold om hvor valide og reliable resultatene er. Ved masterprosjektets ende er antall mulige poeng ujevnt fordelt mellom ferdighetstestens tre kategorier og de tilhørende delferdighetene, og dette er en utfordring med hensyn til testens innholdsvaliditet. Testen som helhet er intern konsistent, og verdien på testens standardfeil er under ett poeng, som antyder at testscore er et reliabelt mål. Likevel er det akseptabel intern konsistens i kun to av tre testkategorier, og to av åtte delferdigheter, som antyder at testen ikke gir reliable scorere på kategori- eller delferdighetsnivå. I tillegg er det lav gjennomsnittlig vanskelighetsgrad i testen, som betyr at testen egner seg best til å differensiere elever på lavt til middels ferdighetsnivå. Validiteten i testen er ikke testet med kvantitative analyser, men kun gjennom en skjønnsmessig validitetsvurdering. Det er derfor behov for en mer systematisk undersøkelse av validiteten til ferdighetstesten.

Testens resultater (FS2) gir innsikt i vanskelighetsgraden til delferdighetene i testen (FS2a). Alle de fire delferdighetene som er knyttet til kategorien «å kunne trekke velgrunnede konklusjoner» er blant de fem delferdighetene med høyest vanskelighetsgrad. I likhet med Kluge et al. (2018) hadde delferdighetene som var knyttet til kategorien «forståelse for naturvitenskapelig informasjon» lav vanskelighetsgrad. De to delferdighetene knyttet til kategorien «strategier for kritisk tenkning» har svært ulik plassering i rangeringen av vanskelighetsgraden til delferdighetene i testen. Det å begrunne og gjøre rede for egne strategier (S1) var den nest vanskeligste delferdigheten i testen, med utgangspunkt i resultatene fra studie 4, mens den delferdigheten med lavest vanskelighetsgrad var tilknyttet oppgaver hvor elevene skulle vurdere troverdigheten til to tekster fra internett (S2).

Testen viser at den kan skille mellom enkelte elevgrupper, med utgangspunkt i de bakgrunnsvariablene som elevene svarte på før de begynte på ferdighetstesten (FS2b). Variabler som ga signifikant forskjell i gjennomsnitt på testscore var om elevene hadde minoritetsspråklig bakgrunn ( $p < 0,01$ ,  $d = 1,07$ ), hvilken karakter de har i naturfag ( $p < 0,01$ ,  $d = 0,83$ ) og karakteren i norsk ( $p < 0,01$ ,  $d = 0,78$ ). Det var også signifikant forskjell i gjennomsnittet til de elevene som trodde at kjøtt var like kreftfremkallende som å røyke før de begynte på testen, og de som ikke trodde dette ( $p < 0,01$ ,  $d = 1,5$ ).

Elevenes kompetanse til å tenke kritisk er avhengig av både kognitive ferdigheter og holdninger. Det er en svak, men signifikant korrelasjon ( $r_s = 0,33$ ,  $p = 0,01$ ) mellom testscore og den sammensatte kategorien «holdninger knyttet til kritisk tenkning», men det er betraktelig variasjon blant elevene innad i de ulike ferdighetsnivåene. Etersom holdningskategoriene ikke har gjennomgått samme kvalitetssikring som ferdighetstesten er det mer usikkerhet knyttet til målingen av dette.

Hvor stor overføringsverdi har resultatene fra studiene? Resultatene fra dette masterprosjektet kan ikke generaliseres til å være gjeldende for ungdomsskoleelever som populasjon, ettersom det ikke er anvendt et tilfeldig utvalg. Resultatene kan likevel tolkes som indikasjoner som bør undersøkes nærmere. Kvaliteten på ferdighetstesten har blitt undersøkt og revidert gjennom fire studier med fire forskjellige utvalg, og dette har gitt innsikt i elevenes kritiske tenkning. Det fjerde og siste studiet, var stort nok til å anvendes som et pilotstudier (Cohen et al., 2007). Med utgangspunkt i dette kan jeg konkludere at det er grunn til å tro at resultatene kan være gjeldene for lignende elevgrupper, selv om jeg ikke kan generalisere.

## 6.1 Videre forskning

Et kartleggingsverktøy for kritisk tenkning har et stort potensial. Ferdighetstesten bør derfor videreutvikles med utgangspunkt i de forslagene som har blitt presentert i forbindelse med studie 4, og det bør undersøkes hvor effektive endringene er. Hvis ferdighetstesten kan kartlegge elevenes ferdigheter til å tenke kritisk på et kategori- og delferdighetsnivå med presise mål kan dette bidra til å gi økt innsikt i elevers kritiske tenkning. Videre kan det utformes flere oppgavesett med utgangspunkt i flere autentiske tekster og problemstillinger. Det vil gi bredere innsikt i den faktiske vanskelighetsgraden til de ulike delferdighetene, og muliggjøre enda strengere utvelgelse av oppgaver med hensyn til item-analyse. For å få bedre innsikt i validiteten til ferdighetstesten bør den undersøkes med faktoranalyse. Validiteten bør også undersøkes ved å teste samme utvalg med elever med en anerkjent og utprøvd test i kritisk tenkning, og med ferdighetstesten, og sammenligne resultatene, for eksempel testen som ble utviklet gjennom PROBE-prosjektet.

For å oppnå kartleggingsverktøyets potensial bør holdningsvariablene videreutvikles og kvalitetssikres eller erstattes. Det bør derfor gjennomføres egne undersøkelser med mange variabler som er ment for å undersøke holdninger til kritisk tenkning generelt og mer temaspesifikke variabler innen naturfag. På den måten kan de variablene som gir mest valide og reliable resultater anvendes i kartleggingsverktøyet. Hvis både holdninger og ferdighetene kan måles, og gi resultater av høy kvalitet, kan elevenes samlede kompetanse til å tenke kritisk undersøkes.

Ettersom undersøkelsen har blitt gjennomført på elever på én skole, er det begrenset hva resultatene kan fortelle om ungdomsskoleelever generelt. Derfor bør kartleggingsverktøyet undersøkes med et større og randomisert utvalg. Først da er det mulig å vite hvilke faktorer som påvirker elevenes kritiske tenkning og ikke. Det er også potensiale for å undersøke hvor mye elevenes kritiske tenkning i naturfag utvikler seg ved å undersøke elever ved videregående skole.



## 6.2 Implikasjoner for undervisning

Elevene må kunne tenke kritisk om naturvitenskapelige problemstillinger, og god veiledning er avhengig av innsikt i elevenes kompetanse. Mitt mål er at naturfagslærere skal kunne anvende kartleggingsverktøyet for få oversikt over hvilke ferdigheter elevene mestrer, og hvilke som bør utvikles videre. Dette muliggjør formativ vurdering, og at naturfagslærere kan tilpasse sin undervisning til elevenes ferdighetsnivå. Et standardisert kartleggingsverktøy som dette bør likevel kun anses som et supplement til lærerens inntrykk av elevenes kompetanse, og ikke en erstatning.

Problemstillingene i ferdighetstesten kan også være et utgangspunkt for undervisning i kritisk tenkning. Det er mulig å diskutere og modellere ulike strategier for å vurdere troverdigheten til de to tekstene i del 1 av testen og hvordan man kan tenke kritisk på del 2 for å finne den rette konklusjonen om hvorfor det er blitt mindre tang og tare på havbunnen utenfor Alaska.

Mitt håp er at masterprosjektet kan bidra med å forberede fremtidige elever på ungdomstrinnet til å tenke kritisk i naturfag og livet for øvrig. Vi lever i en spesiell tid i verdenshistorien, og den makten til å påvirke verden som teknologisk og vitenskapelig utvikling har gitt oss, krever borgere som kan ta gode, gjennomtenkte og velbegrunnede avgjørelser for både seg selv, samfunnet og alt liv på planeten.

## Litteraturliste

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R. & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of educational research*, 78(4), 1102-1134.
- Aidley, D. (2018). *Introducing Quantitative Methods: A Practical Guide*. Macmillan Education UK.
- Akerbæk, E., Karlsen, M. L. & Skiphamn, S. S. (2017, 15.09.2017). *Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke*. Faktisk.no. Hentet 03.09.2020 fra <https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>
- AssesmentDay. *Watson Glaser Critical Thinking Appraisal*. Mercury House. Hentet 11.10.20 fra <https://www.assessmentday.co.uk/watson-glaser-critical-thinking.htm>
- Bailin, S. (2002). Critical thinking and science education. *Science & Education*, 11(4), 361-375.
- Bichi, A. (2016). Classical test theory: An Introduction to Linear Modeling Approach to Test and Item Analysis. *International Journal for Social Studies*, 2, 27-33.
- Bjørndal, A. & Hofoss, d. (2004). *Statistikk for helse- og sosialfagene* (2. utg.). Gyldendal Akademisk.
- Brodin, E. (2007). *Critical thinking in scholarship: meanings, conditions and development*. Lund University.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage publications.
- Cohen, L., Manion, L. & Morrison, K. R. B. (2007). *Research Methods in Education*. Routledge.
- Cronbach, L. J. & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, 64(3), 391-418.
- Dahl, T., Askling, B., Hegge, K., Kulbrandstad, L., Lauvdal, T., Qvotrup, L., Salvanes, K., Skrøvseth, S., Thue, F. & Mausethagen, S. (2016). *Ekspertgruppa om lærerrollen Om lærerrollen: et kunnskapsgrunnlag*. Bergen: Fagbokforlag.

- Dewey, J. (1910). *How We Think* (1997. utg.). Dover Publications.
- Downing, S. (2006). Selected Response Item Formats in Test Development. I S. H. Downing, T. M. (Red.), *Handbook of Test Development*. Lawrence Erlbaum Associates Inc.
- Eikemo, T. A. & Clausen, T. H. (2012). *Kvantitativ analyse med SPSS: en praktisk innføring i kvantitative analyseteknikker*. Tapir akademisk forl.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership*, 43(2), 44-48.
- Ennis, R. H. (1987). Critical thinking and the curriculum. *Thinking skills instruction: Concepts and techniques*, 40-48.
- Eriksen, T. H. (2018, 22.06.2018). Hvem kan vi stole på? Morgenbladet, Del 1. *Morgenbladet*, s. 26-29.
- Eriksson, L. T. & Blomgren, E. (2020). *Kritisk tenkning* (1. utgave. utg.). Gyldendal.
- Estes, J. A., Tinker, M. T., Williams, T. M. & Doak, D. F. (1998). Killer whale predation on sea otters linking oceanic and nearshore ecosystems. *science*, 282(5388), 473-476.
- Facione, P. A. (1990a). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report).
- Facione, P. A. (1990b). The California Critical Thinking Skills Test--College Level. Technical Report# 1. Experimental Validation and Content Validity.
- Facione, P. A. (2000). The disposition toward critical thinking: its character, measurement, and relationship to critical thinking skill. . *Informal Logic*, Vol 20, No 1 (2000).
- Facione, P. A. (2015). Critical thinking: What it is and why it counts. *Insight assessment*, 2007(1), 1-23.
- Facione, P. A. (2018). Peter Facione - The educational power of measuring Critical Thinking [Forelesning]. Hentet 11.12.2020, fra <https://www.youtube.com/watch?v=xiYQFhg5r24>
- Facione, P. A., Facione, N. C. & Giancarlo, C. A. F. (2000). *The California critical thinking disposition inventory: CCTDI test manual*. California Acad. Press.
- Farmer, C. A., Kaat, A. J., Thurm, A., Anselm, I., Akshoomoff, N., Bennett, A., Berry, L., Bruchey, A., Barshop, B. A. & Berry-Kravis, E. (2020). Person ability scores as an

- alternative to norm-referenced scores as outcome measures in studies of neurodevelopmental disorders. *American Journal on Intellectual and Developmental Disabilities*, 125(6), 475-480.
- Ferguson, L. E. & Bråten, I. (2013). Student profiles of knowledge and epistemic beliefs: Changes and relations to multiple-text comprehension. *Learning and Instruction*, 25, 49-61.
- Ferguson, L. E. & Krange, I. (2020). Hvordan fremme kritisk tenkning i grunnskolen? *Norsk pedagogisk tidsskrift*, 104(02), 194-205.
- Ferrer, M. & Wetlesen, A. (2019). Kritisk tenkning i samfunnsfag. I *Kritisk tenkning i samfunnsfag*. Universitetsforlaget.
- Finkelstein, L. (1982). What is not measurable, make measurable. *Measurement and Control*, 15(1), 1.
- Fisher, A. (2011). *Critical thinking: An introduction*. Cambridge university press.
- Friborg, O. (2010). Klassisk testteori og utvikling av spørreinstrumenter. I M. R. Martinussen, D. Arai, O. Friborg, K. A. Hagtvet, B. Handegård, J. H., B. K., S. Lie & W.-T. Mørck (Red.), *Kvantitativ forskningsmetodologi i samfunns- og helsefag*. Fagbokforlaget.
- Frønes, T. (2017). Å lese og navigere på nettet. En studie av elevers navigasjonsstrategier. *Dr. art.-avhandling, Universitetet i Oslo*.
- Gaukroger, S. (1995). *Descartes : an intellectual biography*. Clarendon Press ; Oxford University Press.
- Gigerenzer, G., Todd, P. M. & Group, A. R. (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press.
- Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, 15(3), 309-333.
- Hansen, A. (2016). *Hjärnastark*. Fitnessförlaget.
- Hansen, H. V. (2002). The Straw Thing of Fallacy Theory: The Standard Definition of 'Fallacy'. *Argumentation*, 16(2), 133-155.
- Hope, K. & Klausen, M. H. (2012). *Skepsis: guide til kritisk tenkning*. Humanist.

- Isachsen, S. E. (2017, 10. september). Åpne øynene. *Sophie Elise*.  
[https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html)
- Kleven, T. A. & Hjordemaal, F. (2018). *Innføring i pedagogisk forskningsmetode: en hjelp til kritisk tolking og vurdering*. Fagbokforlaget.
- Kluge, A., Dolonen, J. A., Scherer, R. & Ludvigsen, S. R. (2018). *Testing and Teaching Critical Thinking and Problem Solving in a Changing World: Summative report from the PROBE-project: Critical thinking and problem-solving skills in a changing world—An approach to the design of classroom-based assessment and teaching*. University of Oslo, Norway.
- Kolstø, S. D. (2006). Et allmenndannende naturfag. Fagets betydning for demokratisk deltakelse. *Nordic Studies in Science Education*, 2(3), 82-99.
- Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking skills and creativity*, 4(1), 70-76.
- Kubinger, K. D., Holocher-Ertl, S., Reif, M., Hohensinn, C. & Frebort, M. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a two solutions and three distractors item format to a one solution and five distractors item format. *International Journal of Selection and Assessment*, 18(1), 111-115.
- Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with the California Critical Thinking Tests. *Nurse Educator*, 22(5), 29-33.
- Lie, S., Hopfenbeck, T. N., Ibsen, E. & Turmo, A. (2005). Nasjonale prøver på ny prøve: rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005. *Acta didactica*
- Loo, R. & Thorpe, K. (1999). A psychometric investigation of scores on the Watson-Glaser critical thinking appraisal new Form S. *Educational and Psychological Measurement*, 59(6), 995-1003.
- Lukhele, R., Thissen, D. & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, 31(3), 234-250.
- Mapeala, R. & Siew, N. M. (2015). The development and validation of a test of science critical thinking for fifth graders. *SpringerPlus*, 4(1), 1-13.

- McAlpine, M. (2002). *A summary of methods of Item Analysis*. The Computer Assisted Assessment Centre.
- McPeck, J. E. (1985). Critical Thinking and the 'Trivial Pursuit' Theory of Knowledge. *Teaching Philosophy*, 8(4), 295-308.
- McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational researcher*, 19(4), 10-12.
- Messick, S. (1993). Trait equivalence as construct validity of score interpretation across multiple methods of measurement. *Construction versus choice in cognitive measurement*, 61-74.
- Millar, R. & Osborne, J. (1998). *Science education for the future*. Nuffield Foundation. K. s. C. London.
- Mork, S. (2008a). Hvordan tilrettelegge for argumentasjon i undervisningen? *Naturfag*, 3(3 2008), 14-16.
- Mork, S. (2008b). Hvorfor argumentasjon i naturfag? *Naturfag*, 3(3 2008), 10-13.
- Nardi, P. M. (2018). *Doing Survey Research: A Guide to Quantitative Methods*. Taylor & Francis.
- NSD. (u.å.-a). *Barnehage- og skoleforskning*.  
<https://www.nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/barnehage-og-skoleforskning/>
- NSD. (u.å.-b). *Hvordan gjennomføre et prosjekt uten å behandle personopplysninger?*  
<https://www.nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/hvordan-gjennomfore-et-prosjekt-uten-a-behandle-personopplysninger/>
- NTB. (2020, 15.10.2020). Mange norske elever sliter med kildekritikk. *Verdens Gang*.  
<https://www.vg.no/nyheter/innenriks/i/Ald5n3/mange-norske-elever-sliter-med-kildekritikk>
- O'Hare, L. (2005). *Measuring critical thinking skills and dispositions in undergraduate students* [Doktorgradsavhandling, Queen's University of Belfast].
- Pallant, J. (2016). *SPSS Survival Manual: a step by step guide to data analysis using IBM SPSS (6th ed. utg.)*. McGraw-Hill.

- Patel, R., Davidson, B. & Larsen, F. B. (1995). *Forskningsmetodikkens grunnlag: å planlegge, gjennomføre og rapportere en undersøkelse*. Universitetsforl.
- Ringdal, K. (2018). *Enhet og mangfold - Samfunnsvitenskapelig forskning og kvantitativ metode* (4. utgave. utg.). Fagbokforlaget.
- Sakariassen, H., Hovden, J. & Moe, H. (2017). Bruksmønstre for digitale nyheter. *Reuters Institute Digital News Report*. Bergen: University of Bergen.
- Schleicher, A. (2019). PISA 2018: Insights and Interpretations. *OECD Publishing*.
- Sirnes, S. M. (2005). *Flervalgsoppgaver: konstruksjon og analyse*. Fagbokforlaget.
- Sjøberg, S. (1986). *Elever og lærere sier sin mening. Rapport fra SISS-prosjektet*. Universitetsforlaget.
- Sjøberg, S. (2009). *Naturfag som allmenndannelse : en kritisk fagdidaktikk* (3. utg. utg. utg.). Gyldendal akademisk. .
- Stensland, M. (2020, 15.10.2020). Å være kritisk til kilder er viktigere enn noen gang. Men norske elever lærer ikke nok om det. *Aftenposten*.  
<https://www.aftenposten.no/norge/i/JJoLWj/aa-vaere-kritisk-til-kilder-er-viktigere-enn-noen-gang-men-norske-eleve>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273-1296.
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53.
- Thagaard, T. (2013). *Systematikk og innlevelse: en innføring i kvalitativ metode* (4. utg. utg.). Fagbokforlag.
- Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J. & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663-682.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.
- Utdanningsdirektoratet. (2019a). *Grunnleggende ferdigheter (naturfag)*. Utdanningsdirektoratet. <https://www.udir.no/lk20/nat01-04/om-faget/grunnleggende-ferdigheter>

- Utdanningsdirektoratet. (2019b). *Overordnet del - Verdier og prinsipper for grunnopplæringen*. Utdanningsdirektoratet. <https://www.udir.no/lk20/overordnet-del/>
- Utdanningsdirektoratet. (2019c). *Overordnet del: Kritisk tenkning og etisk bevissthet*. . Utdanningsdirektoratet. <https://www.udir.no/lk20/overordnetdel/opplaringens-verdigrunnlag/1.3-kritisk-tenkning-og-etisk-bevissthet/>
- Utdanningsdirektoratet. (u.å.). *Rammeverk for nasjonale prøver*. Utdanningsdirektoratet. <https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover/hva-er-nasjonale-prover/#>
- Vieira, R. M. & Tenreiro-Vieira, C. (2016). Fostering scientific literacy and critical thinking in elementary science education. *International Journal of Science and Mathematics Education*, 14(4), 659-680.
- Vieira, R. M., Tenreiro-Vieira, C. & Martins, I. P. (2011). Critical thinking: Conceptual clarification and its importance in science education. *Science education international*, 22(1), 43-54.
- Walton, D. (2005). *Fundamentals of critical argumentation*. Cambridge University Press.
- Walton, D. (2010). Why fallacies appear to be better arguments than they are. *Informal Logic*, 30(2), 159-184.
- Weyergang, C. & Frønes, T. S. (2020). Kapittel 7. Å lese kritisk: Elevers vurderinger av teksters troverdighet og pålitelighet. I T. S. F. F. Jensen (Red.), *Like muligheter til god leseforståelse?* (s. 166-195).
- Wibowo, S. B. (2016). *Misuses Cronbach Alpha On Achievement Tests*. ICERE 2016, Indonesia.



## Vedlegg

### Liste over vedleggene

#### 1. Vedlegg til studie 1:

- 1.1. Observasjonsskjema anvendt i studie 1
- 1.2. Test anvendt i studie 1 (del 1 og 2)
- 1.3. Test anvendt i studie 1 (del 3)

#### 2. Vedlegg til studie 2:

- 2.1. Test anvendt i studie 2: Venneutvalget
- 2.2. Test anvendt i studie 2: Lærerutvalget

#### 3. Vedlegg til studie 3:

- 3.1. Kartleggingsverktøy anvendt i studie 3
- 3.2. Resultater på reliabilitetsanalyse av holdningskategoriene i studie 3

#### 4. Vedlegg til studie 4:

- 4.1. Oversikt over kategoriene, delferdighetene og oppgavene i studie 4
- 4.2. Løsningsforslag og poengutdeling på testens åpne oppgaver
- 4.3. Observasjonsskjema til faglærerne i studie 4
- 4.4. Kartleggingsverktøy anvendt i studie 4
- 4.5. Resultater reliabilitetsanalyse på holdningskategoriene i studie 4

## Vedlegg til studie 1:

### Vedlegg 1.1: Observasjonsskjema anvendt i studie 1

<b>Del 1</b>  <b>Arguments- oppgaver</b>	Spørsmål om oppgaveformulering:
	Spørsmål om begreper og faginnhold
<b>Del 2</b>  <b>Konklusjoner</b>	Spørsmål om oppgaveformulering
	Spørsmål om begreper og faginnhold
<b>Del 3</b>  <b>Kildekritikk</b>	Spørsmål om oppgaveformulering
	Spørsmål om kildene
	Spørsmål om begreper og faginnhold

Øvrige notater:

## Vedlegg 1.2: Test anvendt i studie 1 (del 1 og 2)

### Informasjon før testen:

**Viktig:** Følg oppgavene fra testen i tekstheftet. Du kan bla tilbake for å finne informasjon, men det er veldig viktig å ikke bla fremover i heftet.

### Informasjon om argumenter

Når man kommer med en påstand, vil man ofte også ha argumenter for å overbevise om at påstanden er sann eller sannsynlig. I et godt argument er det ofte en begrunnelse. Det er selve beviset i argumentet. Se på de to eksemplene under:

«Jeg mener at det er galt å fiske med krok, fordi vi vet at mange dyr kan føle smerte, og fisk er et dyr.»

«Du burde ikke stupe fra fjellet og ut i havet, for det gjorde broren min, og han brakk nesa.»

Hele setningen er argumentet, påstanden er markert i gult og begrunnelsen i grønt.

### Del 1: Argumenter

Når man skal ta avgjørelser er det viktig å kunne skille mellom sterke og svake argumenter.

Du får nå en kort tekst med informasjon, og et spørsmål. Til dette vil du bli presentert for argumenter som du skal vurdere. Du kan enten svare at det er et sterkt eller svakt argument, ut fra kriteriene under.

Et sterkt argument kjennetegnes som:

- Direkte relevant til spørsmålet
- Bruker viktig og konkret informasjon

Et svakt argument kjennetegnes ved at det oppfyller et av disse kriteriene:

- Ikke direkte relevant for å svare på spørsmålet
- Bruker ikke viktig og konkret informasjon.
- Uttrykker en mening, og ikke fakta.

Ta utgangspunkt i at informasjonen i teksten og argumentene er sann, og husk at du bare skal vurdere argumentene, og ikke uttrykke din mening om saken.

### Del 1.1: Solcellepanel

Ola har et stort hus, og han lurer på om han skal kjøpe solcellepanel til å ha på taket. Et solcellepanel er en energikilde som samler solenergi fra sola, og gjør den om til elektrisk energi som Ola kan bruke i huset. Da trenger ikke Ola å kjøpe strøm fra strømselskapene.

Bør Ola kjøpe solcellepanel for å spare penger?

1. Ja, det bør han
  - Sterkt argument
  - Svakt argument
2. Ja, for solcellepanel er en klimavennlig energikilde.
  - Sterkt argument
  - Svakt argument
3. Ja, for de som har solcellepanel på taket har gjennomsnittlig bedre råd.
  - Sterkt argument
  - Svakt argument
4. Nei, han vil ikke spare penger på å kjøpe solcellepanel fordi solcellepanel er dyrt, og har lav gjennomsnittlig levetid. Den går altså ofte i stykker før man har tjent inn det solcellepanelene kostet.
  - Sterkt argument
  - Svakt argument
5. Ja, solcellepanel er en god investering. I løpet av fem år vil han ha spart like mye penger på strømgifter som han brukte på å kjøpe solcellepanelene. Etter det vil han fortsette å spare penger.
  - Sterkt argument
  - Svakt argument
6. Ja, banken til Ola sier at solcellepanel er en god investering
  - Sterkt argument
  - Svakt argument
7. Ja, mange av som har solcellepanel er veldig fornøyde med dem
  - Sterkt argument
  - Svakt argument

### *Del 1.2: Genteknologi*

I genene våre står oppskriften til hvert enkelt menneske. Med fremtidens genteknologi kan vi sannsynligvis forandre oppskriften til levende organismer som bakterier, planter og dyr. Med genteknologien vil fremtidens leger kanskje kunne påvirke og forandre egenskaper hos et barn, som for eksempel øyefarge eller hvilket kjønn et barn får. Noen sykdommer blir man født med og mange av disse kan kureres ved hjelp av genteknologi.

Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke?

1. Nei, det synes jeg ikke.
  - Sterkt argument
  - Svakt argument
2. Ja, ved å bruke genteknologi til å kurere sykdommer vil det bli færre syke i samfunnet, og man kan bruke tiden på å behandle andre syke pasienter på sykehusene.
  - Sterkt argument
  - Svakt argument
3. Ja, man bør alltid jobbe for å forbedre naturen. Samfunnet vil bli bedre fordi genteknologi kan gjøre at mennesker blir bedre på jobbene sine.
  - Sterkt argument
  - Svakt argument

4. Nei, å forandre på naturen er ikke riktig. Hvis man begynner å forandre på oppskriften til dyr og mennesker, vil det bare være et spørsmål om tid før det ikke er noen grenser på hvilke endringer vi har gjort på mennesker.
  - Sterkt argument
  - Svakt argument
5. Nei, den genetiske oppskriften til mennesker er komplisert, hvis vi endrer på en egenskap vil vi i mange tilfeller også endre på andre egenskaper hos mennesker, som vi ikke vil endre på.
  - Sterkt argument
  - Svakt argument
6. Ja, en kjent professor som heter Audun Smart ser ingen utfordringer med å bruke genteknologi for å helbrede syke.
  - Sterkt argument
  - Svakt argument
7. Nei, for enten må vi tillate at vi kan forandre på alle egenskaper hos mennesket, eller så må vi forby det. Å kun helbrede syke er ikke mulig.
  - Sterkt argument
  - Svakt argument

### *Del 1.3: Åpen oppgave:*

Hvordan tenkte du når du skulle vurdere argumentene over? Er det noen av argumentene som var spesielt vanskelige å vurdere? Hvorfor var de vanskelige?

### *Del 2: Hva er forklaringen?*

I denne serien av oppgaver vil du stegvis bli presentert for ny informasjon om hvorfor tang og tare forsvinner fra havbunnen utenfor Alaska. Målet er at du ut fra denne informasjonen skal finne den mest sannsynlige forklaringen på fenomenet.

#### *Del 2.1: Bruk informasjonen, og vurder forklaringen/konklusjonen:*

- Tang og tare er sjøplanter som lever på bunnen av havet.
- I havbunnen på utsiden av Alaska har det blitt observert over flere år at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt på grunn av global oppvarming.

#### **Forklaring:**

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur

a) Er du enig i denne konklusjonen kan stemme?

- Ja
- Nei

b) Hvilke andre alternative forklaringer kan du se for deg?

- c) Er det tilstrekkelig med informasjon for å trekke en konklusjon nå? Hvis ikke hva slags informasjon mangler for at du skal være sikker på hva som er forklaringen på at det har blitt mindre tang og tare på havbunnen utenfor Alaska?

### ***Del 2.2: Hvilke forklaringer er mulige?***

*Bruk informasjonen fra forrige deloppgave og informasjonen under til svare på spørsmål om årsaken om hvorfor det har blitt mindre tang og tare på havbunnen.*

- Mengden kråkeboller har gått opp i samme tidsrom som det har blitt mindre tang og tare.
  - Sjøotere er et rovdyr, og spiser kråkeboller.
  - Kråkebollen er et byttedyr.
  - Tang og tare er tilhører planteriket, og er ikke et dyr.
  - Tidligere var det et problem mennesker jaktet på sjøotere for å bruke pelsen deres til å lage eksklusive pelsprodukter som pelskåper. Dette førte til at sjøoterne nesten ble utryddet i Alaska på starten av 1900-tallet.
  - Det ble ulovlig å jakte på sjøotere noen år senere fordi sjøoterne var på vei til å bli utryddet.
  - I 1970 var det like mange sjøotere i Alaska som det var før menneskene begynte å jakte på den. Å forby jakt på sjøotere hadde vært en suksess.
  - Spekkhoggere er en hval, som lever av å jakte på mindre byttedyr, og spiser for det meste fisk.
  - Et økosystem består av mange arter (altså ulike typer dyr, planter og bakterier), som påvirker hverandre.
  - Et byttedyr er maten for et rovdyr. Hvis byttedyret forsvinner, vil rovdyret enten dø av sult eller måtte spise en annen art.
  - Mange arter er sårbare for endringer i økosystemet, som forandringer i temperatur eller hvis det kommer nye ukjente arter inn i økosystemet.
  - Mennesker påvirker økosystemer blant annet gjennom global oppvarming, forsøpling av havene og ved å overfiske.
  - Hvis mennesker fortsetter å kaste søppel i havene vil det i 2050 være mer plastikk i havet enn fisk.
  - Sportsfiskere klager på at det har blitt vanskeligere å få fisk når de fisker.
- a) Hva tenker du er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen?

- b) Mangler du informasjon for å være sikker på hva som er forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? Hva slags informasjon mangler du i så fall for å være sikker på konklusjonen din?

- c) Vurder de mulige forklaringene under for hvorfor det har blitt mindre tang og tare på havbunnen:
1. Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.
  2. Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.
  3. Plastikk fører til at fiskebestanden har gått ned.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.
  4. Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.
  5. Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.
  6. Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
    - Dette er en mulig forklaring.
    - Dette er en ikke en mulig forklaring.

### *Del 2.3: Hvilken forklaring er den mest sannsynlige?*

Bruk informasjonen fra de forrige deloppgavene og informasjonen under til svare på spørsmål om årsaken om hvorfor det har blitt mindre tang og tare på havbunnen.

- En spekkhogger kan spise 1825 sjøotere i året hvis den bare spiser otere.
- Man kan finne tang og tare i områder syd for Alaska, hvor det er et varmere klima.
- Det har blitt 25% færre sjøotere i året i Alaska siden 1990-tallet.
- Forskerne er forundret over at de ikke har funnet flere døde sjøotere på strender. Dette gjør at de utelukker sykdom som hovedårsaken til hvorfor det har blitt færre sjøotere, fordi syke sjøotere ville blitt skylt opp på strendene når de døde.

- a) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen?
- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at fiskebestanden har gått ned.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

b) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen?



## Vedlegg 1.3: Test anvendt i studie 1 (del 3)

### *Informasjon før testen:*

**Viktig:** Følg oppgavene fra testen i tekstheftet. Du kan bla tilbake for å finne informasjon, men det er veldig viktig å ikke bla fremover i heftet.

### **Informasjon om argumenter**

Når man kommer med en påstand, vil man ofte også ha argumenter for å overbevise om at påstanden er sann eller sannsynlig. I et godt argument er det ofte en begrunnelse. Det er selve beviset i argumentet. Se på de to eksemplene under:

«Jeg mener at det er galt å fiske med krok, fordi vi vet at mange dyr kan føle smerte, og fisk er et dyr.»

«Du burde ikke stupe fra fjellet og ut i havet, for det gjorde broren min, og han brakk nesa.»

Hele setningen er argumentet, påstanden er markert i gult og begrunnelsen i grønt.

### *Del 3: Kildekritikk*

Denne delen av testen er tredelt. I første og andre del skal du lese et utdrag fra en kilde og svare på spørsmål knyttet til kildene. I den tredje delen skal du sammenligne kildene.

#### *Del 3.1: Analyse av kilde 1: «Åpne øynene»*

«For noen måneder siden så jeg en dokumentar som satt meg ut. "What the health" heter den, og ligger på Netflix. Den provoserte meg, men det er helt ærlig nesten da jeg liker meg selv best. Når jeg er engasjert i noe på denne måten. Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer, og jeg har SÅ lyst at dere skal se den. Men først, kan jeg komme med et par små fakta som jeg noterte meg underveis i filmen:

– På et slakteri slaktes 250 kuer i timen. Dette gir 15 sekunder per ku. Ting går fort, noe som har ført til at det blant annet er spør av avføring i 90% av kjøttet som selges i dagligvarebutikker i USA. (norge har – i følge et par artikler jeg fant, halvparten). For å ikke snakke om spor av puss, og all antibiotikaen og hormonene man får i seg ved å spise kjøtt / meieriprodukter. Her er ikke norge noe annerledes.

- *Å spise kjøtt er like kreftfremkallende som å røyke.*
- *Kun 6 – 10% av alle krefttyper kan overføres genetisk. Om så, så er det ikke en eller to kreftceller som er ille, men maten du spiser kan, og VIL booste disse cellene om du spiser feil (det vil si, for mye kjøtt). Som de fleste av oss gjør..*
- *Egg er såpass lite godt for deg at det nå i amerika ikke engang er lovlig å si i reklamer at det er sunt eller bra for deg med hele setninger. Så de må heller si “a good egg”, det er ikke lovlig å si “good for YOU”.*
- *Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør. Dette fører til at man vil ha mer og mer, uansett om man går opp i vekt og huden tar skade.*
- *Det verste er at alt handler om kun en ting – penger. Apotekene skal tjene penger, regjeringen skal tjene penger, og kjøttindustrien skal. Disse tre jobber ofte sammen. Altså, bare SE den dokumentaren, den er så bra!*



*Det er trist at alt handler om penger. Både våre egne hoder, men samfunnet vi lever i.*

*Ville du ha banket din beste venninne for 1 million? Ville du ha spart på en hemmelighet som gjorde et helt land syke med diabetes, hjertefeil, astma og kreft, om du hvert år tjente 10 millioner? Hva er prisen for våre verdier? Har moral en prislapp, og i så tilfelle – er da våre moraler egentlig ekte? Slik funker denne verden.. og det finnes kun en ting å gjøre: åpne øynene og ikke tro du ikke blir påvirket, still spørsmål og vær din makt bevist. Innser du det, kan du ta steget videre og GJØRE noe. Om ikke for miljøet, om ikke for dyrene, men for din egen del.»*

Hentet 03.09.2020 fra innlegget «Åpne øynene», Sophie Elises blogg, 10. september 2017 ([https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html))

### **Oppgaver til kilde 1:**

- A) Oppsummer hva forfatterens **budskap** (påstanden) er i teksten, altså hva er det forfatteren prøver å si med teksten?

- B) Oppsummer hva forfatterens viktigste **Begrunnelser** (bevis) for budskapet i teksten.

- C) Hvordan vil du beskrive teksten som kilde, og hvor troverdig synes du den er? Begrunn hvorfor.

D) Var informasjonen i kilde 1 troverdig?

- Ja
- Nei

### **Del 3.2: Kilde 2 «– Ikke like kreftfremkallende»**

«Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning. Direktør Giske Ursin bekrefter at det **ikke er like kreftfremkallende** å spise kjøtt som å røyke. Ursin tror hun vet hvor misforståelsen om at det er like kreftfremkallende å spise kjøtt som å røyke kommer fra:

I 2015 klassifiserte Verdens helseorganisasjon (WHO) prosessert kjøtt (som for eksempel pølser) som sikkert kreftfremkallende, mens rødt kjøtt er klassifisert som sannsynlig kreftfremkallende. Dermed kommer prosessert kjøtt nå i samme klasse som røyk.



At både prosessert kjøtt og røyk er klassifisert som sikkert kreftfremkallende, er ikke det samme som at risikoen er like stor. Graden av kreftfremkallende er ekstremt forskjellig.

Rødt, saltet og bearbeidet kjøtt kan **øke risikoen** for tarmkreft. Ifølge Kreftregisteret er det imidlertid bare en moderat andel av forekomsten av tarmkreft som kan tilskrives kosthold og livsstil. Ursin sier følgende: - For lungekreft kan derimot over 80 prosent tilskrives røyking.

Direktøren understreker at Kreftregisteret støtter kreftrådene fra WHO som anbefaler å begrense inntaket av rødt og bearbeidet kjøtt. Dette fordi noe av kreftforekomsten kan forebygges med godt kosthold og en aktiv livsstil.»

Hentet 03.09.2020 fra «Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke», Faktisk.no, 15.09.2017 (<https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>)

### **Oppgaver til kilde 2:**

A) Oppsummer hva forfatterens budskap (påstanden) er i teksten, og de viktigste begrunnelsene (bevisene) som styrker budskapet

B) Var informasjonen i kilde 2 troverdig?

- Ja
- Nei

***Del 3.3: Oppgaver til begge kildene:***

A) Synes du kilde 1 eller kilde 2 var mest troverdig, og hvorfor synes du det?

B) Hva tenker du om påstanden "Det er like kreftfremkallende å spise kjøtt som å røyke" etter å ha lest begge tekstene?

C) Hva tenker du kjennetegner en troverdig kilde?

## Vedlegg til studie 2:

### Vedlegg 2.1: Test anvendt i studie 2: Venneutvalget

Side 1

## Test av ferdigheter innen kritisk tenkning i naturfag

Denne testen er en del av et masterprosjekt som har som mål å utvikle en god test for kritisk tenkning i naturfag. Testen er anonym, og svarene på testen skal brukes for å videreutvikle testen. Noen viktige tips til testen

- Les oppgavene nøye.
- Du kan bla i testen, men unngå å endre på oppgaver du allerede har besvart.
- For å ikke bruke for mye tid på de oppgavene som krever at du skriver svaret, kan det vært lurt å svare i punkter istedet for hele avsnitt.
- Vær ærlig i besvarelsen din.

På forhånd takk for at du bidrar til å utvikle testen.



Side 2

Svar på spørsmålene under, før du begynner på del 1 av testen.

Tror du det kan være kreftfremkallende å spise kjøtt? \*

Ja

Nei

Tror du det er like kreftfremkallende å spise kjøtt som å røyke? \*

Ja

Nei

## Del 1 av testen: Kilder fra internett

1.0 a) Tenk at du skal være så kildekritisk som mulig når du møter en tekst fra internett. Nevn så mange faktorer du klarer om hva som påvirker om kilden er troverdig eller ikke: \*

Denne delen av testen er tredelt.

- I første og andre del skal du lese et utdrag fra en kilde og svare på spørsmål knyttet til kildene.

- I den tredje delen skal du sammenligne kildene.

## 1.1 Kilde 1

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### "Åpne øynene

For noen måneder siden så jeg en dokumentar som satt meg ut. "What the health" heter den, og ligger på Netflix. Den provoserte meg, men det er helt ærlig nesten da jeg liker meg selv best. Når jeg er engasjert i noe på denne måten. Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer, og jeg har SÅ lyst at dere skal se den. Men først, kan jeg komme med et par små fakta som jeg noterte meg underveis i filmen:

– På et slakteri slaktes 250 kuer i timen. Dette gir 15 sekunder per ku. Ting går fort, noe som har ført til at det blant annet er spør av avføring i 90% av kjøttet som selges i dagligvarebutikker i USA. (norge har – i følge et par artikler jeg fant, halvparten). For å ikke snakke om spor av puss, og all antibiotikaen og hormonene man får i seg ved å spise kjøtt / meieriprodukter. Her er ikke norge noe annerledes.

- Å spise kjøtt er like kreftfremkallende som å røyke.
- Kun 6 – 10% av alle krefttyper kan overføres genetisk. Om så, så er det ikke en eller to kreftceller som er ille, men maten du spiser kan, og VIL booste disse cellene om du spiser feil (det vil si, for mye kjøtt). Som de fleste av oss gjør..
- Egg er såpass lite godt for deg at det nå i amerika ikke engang er lovlig å si i reklamer at det er sunt eller bra for deg med hele setninger. Så de må heller si "a good egg", det er ikke lovlig å si "good for YOU".
- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør. Dette fører til at man vil ha mer og mer, uansett om man går opp i vekt og huden tar skade.
- Det verste er at alt handler om kun en ting – penger. Apotekene skal tjene penger, regjeringen skal tjene penger, og kjøttindustrien skal. Disse tre jobber ofte sammen. Altså, bare SE den dokumentaren, den er så bra!

Det er trist at alt handler om penger. Både våre egne hoder, men samfunnet vi lever i."



Hentet 03.09.2020 fra innlegget «Åpne øynene», Sophie Elises blogg, 10. september 2017 ([https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html))

1.1 a) Var informasjonen i kilde 1 troverdig? \*

- Ja
- Nei

1.1 b) Hvilket av utsagnene om kilde 1 er du mest enig i? \*

- Den er troverdig, fordi forfatteren skriver på en måte jeg forstår.
- Den er troverdig, fordi forfatteren har undersøkt temaet nøye, og har gode begrunnelser.

- Den er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser
- Den er ikke troverdig, fordi alt bloggere skriver er skrevet kun for å selge produkter.

### Informasjon til de neste to oppgavene:

Et argument består av en påstand og en begrunnelse. I eksemplene på argumenter under er påstandene markert i **gult** og begrunnelsene i **grønt**.

«**Jeg mener at det er galt å fiske med krok**, fordi **vi vet at mange dyr kan føle smerte**, og fisk er et dyr.»

«**Du burde ikke stupe fra fjellet og ut i havet**, for **det gjorde broren min, og han brakk nesa**.»

I en tekst er **påstander** budskap i teksten. **Begrunnelser** brukes for å bygge opp troverdigheten til forfatterens budskap. Hele argumenter må i en tekst inneholde både en påstand og en begrunnelse.

#### 1.1 c) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 1?

- "Å spise kjøtt er like kreftfremkallende som å røyke"

- En påstand
- En begrunnelse
- Et helt argument

#### 1.1 d) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 1?

- "For noen måneder siden så jeg en dokumentar som heter "What the health" og ligger på Netflix."

- En påstand
- En begrunnelse
- Et helt argument

#### 1.1 e) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 1?

- "Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør."

- En påstand
- En begrunnelse
- Et helt argument

## 1.1 f) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 1?

- "Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer"
- En påstand
- En begrunnelse
- Et helt argument

## 1.1 g) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

## 1.1 h) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 1 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 1.

- Enig
- Uenig

## 1.1 i) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 1 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

## 1.1 j) Er du enig i hele utsagnet under? \*

Kilde 1 er troverdig, fordi påstandene stemmer med det jeg har lært på skolen om kjøtt og røyking fra før.

- Enig
- Uenig

## 1.1 l) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser.

- Enig
- Uenig

## 1.1 m) Er du enig i hele utsagnet under? \*



Det er ikke nødvendigvis sant at det tar 15 sekunder å slakte en ku, hvis det slaktes 250 kuer i timen på et slakteri.

- Enig
- Uenig

 Sideskift

Side 3

## 1.2 Kilde 2

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### «– Ikke like kreftfremkallende

Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning. Direktør Giske Ursin bekrefter at det ikke er like kreftfremkallende å spise kjøtt som å røyke. Ursin tror hun vet hvor misforståelsen om at det er like kreftfremkallende å spise kjøtt som å røyke kommer fra:



I 2015 klassifiserte Verdens helseorganisasjon (WHO) prosessert kjøtt (som for eksempel pølser) som sikkert kreftfremkallende, mens rødt kjøtt er klassifisert som sannsynlig kreftfremkallende. Dermed kommer prosessert kjøtt nå i samme klasse som røyk. At både prosessert kjøtt og røyk er klassifisert som sikkert kreftfremkallende, er ikke det samme som at risikoen er like stor. Graden av kreftfremkallenhet er ekstremt forskjellig.

Rødt, saltet og bearbeidet kjøtt kan øke risikoen for tarmkreft. Ifølge Kreftregisteret er det imidlertid bare en moderat andel av forekomsten av tarmkreft som kan tilskrives kosthold og livsstil. Ursin sier følgende: - For lungekreft kan derimot over 80 prosent tilskrives røyking. Direktøren understreker at Kreftregisteret støtter kreftrådene fra WHO som anbefaler å begrense inntaket av rødt og bearbeidet kjøtt. Dette fordi noe av kreftforekomsten kan forebygges med godt kosthold og en aktiv livsstil.»

Hentet 03.09.2020 fra «*Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke*», Faktisk.no, 15.09.2017 (<https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>)

1.2 a) Var informasjonen i kilde 2 troverdig? \*

- Ja
- Nei

1.2 b) Hva fortalte kilde 2 om kreftfaren med å spise rødt kjøtt? \*

- Det er like kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt som å røyke.
- Det er ikke kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.

- Det kan være kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det er kun bearbeidet kjøtt som er like kreftfremkallende som å røyke

### 1.2 c) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 2?

- "Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning."
- En påstand
  - En begrunnelse
  - Et helt argument

### 1.2 d) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 2?

- "Rødt, saltet og bearbeidet kjøtt kan øke risikoen for tarmkreft."
- En påstand
  - En begrunnelse
  - Et helt argument

### 1.2 e) \*

Er sitatet under ment som en påstand, en begrunnelse eller et helt argument i kilde 2?

- "Det er ikke like kreftfremkallende å spise kjøtt som å røyke."
- En påstand
  - En begrunnelse
  - Et helt argument

### 1.2 f) Er du enig i hele utsagnet under? \*

Kilde 2 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

### 1.2 g) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 2 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 2.

- Enig
- Uenig

### 1.2 h) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 2 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

### 1.2 i) Er du enig i hele utsagnet under? \*

Påstandene i kilde 2 stemmer med det de fleste tror om kjøtt og røyking fra før, og dette styrker troverdigheten.

- Enig
- Uenig




Side 4

## 1.3. Sammenligning av kilde 1 og 2

### 1.3 a) Hvilken kilde synes du var mest troverdig \*

- Kilde 1 (Sophie Elise «Åpne øynene»)
- Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»)


### 1.3 a.a Hva var den viktigste faktoren når du synes kilde 1 var mest troverdig?

-  Dette elementet vises kun dersom alternativet «Kilde 1 (Sophie Elise «Åpne øynene»» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

Velg kun en av alternativene under.

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

### 1.3 a.b Hva var den viktigste faktoren når du synes kilde 2 var mest troverdig?

-  Dette elementet vises kun dersom alternativet «Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

Velg kun en av alternativene under.

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

1.3 b) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det er like kreftfremkallende å spise kjøtt som å røyke»? \*

- Ja
- Nei

1.3 c) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det kan være kreftfremkallende å spise kjøtt»? \*

- Ja
- Nei



## Del 2 av testen: Vurdering av argumenter

Informasjonen under er viktig når du løser de neste 14 oppgavene.

Når man skal ta avgjørelser er det viktig å kunne skille mellom sterke og svake argumenter. Under finner du kjennetegn for begge.

Et **sterkt** argument kjennetegnes som:

- Direkte relevant til spørsmålet
- Bruker viktig og konkret informasjon

Et **svakt** argument kjennetegnes ved at det oppfyller et eller flere av disse kriteriene:

- Ikke direkte relevant for å svare på spørsmålet
- Uttrykker en mening, og ikke fakta.
- Begrunnelsen kan være relevant for spørsmålet, men det er ikke sammenheng mellom begrunnelsen og påstanden.

I oppgavene **2.1 og 2.2** (under) vil du bli presentert med en kort tekst med informasjon, og et spørsmål. Til dette vil du bli presentert for argumenter som du skal vurdere.

Ta utgangspunkt i at informasjonen i teksten og argumentene er sanne, og husk at du bare skal vurdere argumentene, og ikke uttrykke din mening om saken.

### 2.1. Bør Ola kjøpe solcellepanel for å spare penger?

Ola har et stort hus, og han lurer på om han skal kjøpe solcellepanel til å ha på taket. Et solcellepanel er en energikilde som samler solenergi fra sola, og gjør den om til elektrisk energi som Ola kan bruke i huset. Da trenger ikke Ola å kjøpe strøm fra strømselskapene.

Vurder om argumentene under er svake eller sterke, og velg alternativet som passer best.

### 2.1 a) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Nei, han vil ikke spare penger på å kjøpe solcellepanel, fordi et solcellepanel er dyrt og har lav gjennomsnittlig levetid. Den går altså ofte i stykker før man har tjent inn det solcellepanelene kostet.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 b) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, det mener Per at Ola bør gjøre.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 c) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, for solcellepanel er en klimavennlig energikilde. Det er viktig at vi velger solcellepanel og andre fornybare energikilder for å stoppe global oppvarming.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 d) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, for de som har solcellepanel på taket har gjennomsnittlig bedre råd.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 e) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, banken til Ola sier at solcellepanel er en god investering

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 f) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, solcellepanel er en god investering. I løpet av fem år vil han ha spart like mye penger på strømuttergifter som han brukte på å kjøpe solcellepanelene. Etter det vil han fortsette å spare penger.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 g) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, mange av dem som har solcellepanel er veldig fornøyde med panelene.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

## 2.2. Bør vi bruke genteknologi for å behandle sykdommer?

I genene våre står oppskriften til hvert enkelt menneske. Med fremtidens genteknologi kan vi sannsynligvis forandre oppskriften til levende organismer som bakterier, planter og dyr. Med genteknologien vil fremtidens leger kanskje kunne påvirke og forandre egenskaper hos et barn, som for eksempel øyefarge eller hvilket kjønn et barn får. Noen sykdommer blir man født med og mange av disse kan kureres ved hjelp av genteknologi.

Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke?

Vurder om argumentene under er svake eller sterke, og velg alternativet som passer best.

2.2 a) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Nei, det synes jeg ikke.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 b) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Ja, ved å bruke genteknologi til å kurere sykdommer vil det bli færre syke i samfunnet, og man kan
  - bruke tiden på å behandle andre syke pasienter på sykehusene.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 c) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Ja, man bør alltid jobbe for å forbedre naturen. Samfunnet vil bli bedre, fordi genteknologi kan
  - gjøre at mennesker blir bedre på jobbene sine.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 d) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

Nei, å forandre på naturen er ikke riktig. Hvis man begynner å forandre på den genetiske oppskriften til arter, vil det bare være et spørsmål om tid før det ikke er noen grenser på hvilke

- endringer vi har gjort på mennesker.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 e) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

Nei, den genetiske oppskriften til mennesker er komplisert, hvis vi endrer på en egenskap vil vi i mange tilfeller også endre på andre egenskaper hos mennesker, som vi ikke ønsker å endre på.

- De som blir behandlet med genteknologi kan derfor bli syke av behandlingen.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 f) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

Ja, en kjent professor som heter Øystein Nilsen ser ingen utfordringer med å bruke genteknologi

- for å helbrede syke.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 g) Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:



- Nei, for vi må enten tillate at vi forandrer på alle egenskaper hos mennesker, eller så bør vi forby det.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, selv om begrunnelsen kan være relevant, men er det ikke god nok sammenheng mellom begrunnelsen og påstanden.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.



Side 6

### Del 3 av testen: Å kunne trekke konklusjoner

I denne serien av oppgaver vil du stegvis bli presentert for ny informasjon om hvorfor tang og tare forsvinner fra havbunnen utenfor kysten ved Alaska. Målet er at du ut fra denne informasjonen skal finne den mest sannsynlige forklaringen på hvorfor det skjer.

#### 3.1 Bruk informasjonen, og vurder forklaringen:

- Tang og tare er sjøplanter som lever på bunnen av havet.
- På havbunnen utenfor Alaska har det gjennom flere år blitt observert at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt grunnet global oppvarming.

#### Konklusjon:

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

Konklusjon: Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur. \*

3.1 a) Er du enig i denne konklusjonen kan stemme?

- Ja
- Nei

3.1 b) Er det tilstrekkelig med informasjon for å trekke en konklusjon nå? \*

- Ja
- Nei



Side 7

3.2 Bruk informasjonen fra forrige deloppgave og informasjonen under til svare på spørsmål om årsaken til hvorfor det har blitt mindre tang og tare på havbunnen.

Husk at du kan bla mellom sidene i del 3 for hvis du vil undersøke tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

## Ny informasjon:

- Mengden kråkeboller har gått opp i samme tidsrom som det har blitt mindre tang og tare.
- Sjøotere er et rovdyr, og spiser kun byttedyr.
- Kråkebollen er et byttedyr.
- Tidligere var det et problem mennesker jaktet på sjøotere for å bruke pelsen deres til å lage eksklusive pelsprodukter som pelskåper. Dette førte til at sjøoterne nesten ble utryddet rundt Alaska på starten av 1900-tallet.
- Et byttedyr er maten til rovdyr. Hvis byttedyret forsvinner vil rovdyret enten dø av sult eller måtte spise en annen art.
- Spekkhoggere er en hval, og et rovdyr. Den spiser for det meste fisk. Den har ingen naturlige fiender.

## 3.2 a) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

## 3.2 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

## 3.2 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

## 3.2 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

## 3.2 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



Side 8

3.3 Denne delen av testen er veldig lik 3.2, hvor du skal vurdere hvilke av konklusjonene som kan stemme. Det er de samme forklaringene, men du får mer informasjon som du kan bruke for å finne ut hvorfor det har blitt mindre tang og tare på havbunnen utenfor Alaska

Husk at du kan bla mellom sidene i del 3 for hvis du vil undersøke tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

Ny informasjon:

- Kråkeboller er planteetere.
- Et økosystem består av mange arter (altså ulike typer dyr, planter og bakterier), som påvirker hverandre.
- Det ble ulovlig å jakte på sjøotere i 1950, fordi sjøoterne var på vei til å bli utryddet.
- Mange arter er sårbare for endringer i økosystemet, som forandringer i temperatur eller hvis det kommer nye ukjente arter inn i økosystemet.
- Mennesker påvirker økosystemer blant annet gjennom global oppvarming, forsøpling av havene og ved å overfiske.
- Hvis mennesker fortsetter å kaste søppel i havene vil det i 2050 være mer plastikk i havet enn fisk.
- Tang og tare er tilhører planteriket, og er ikke et dyr.
- Sportsfiskere klager på at det har blitt vanskeligere å få fisk når de fisker.
- I 1970 var det like mange sjøotere i Alaska som det var før menneskene begynte å jakte på den. Å forby jakt på sjøotere har vært en suksess.
- Det er lite ulovlig jakt på sjøotere.

3.3 a) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

3.3 a) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

### 3.4 Denne gangen skal du velge ut den mest sannsynlige forklaringen f or hvorfor det har blitt mindre tang og tare på havbunnen utenfor Alaska

Husk at du kan bla mellom sidene i del 3 for hvis du vil undersøke tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

#### Ny informasjon:

- En spekkhogger kan spise 1825 sjøotere i året, hvis den bare spiser otere.
- Man kan finne samme type tang som finnes på havbunnen utenfor Alaska, i områder lenger syd, hvor det er et varmere klima.
- Det har blitt 25% færre sjøotere i året på havbunnen utenfor Alaska siden 1990-tallet.
- Forskerne er forundret over at de ikke har funnet flere døde sjøotere på strender. Dette gjør at de utelukker sykdom som hovedårsaken til hvorfor det har blitt færre sjøotere, fordi syke sjøotere ville blitt skylt opp på strendene når de døde.

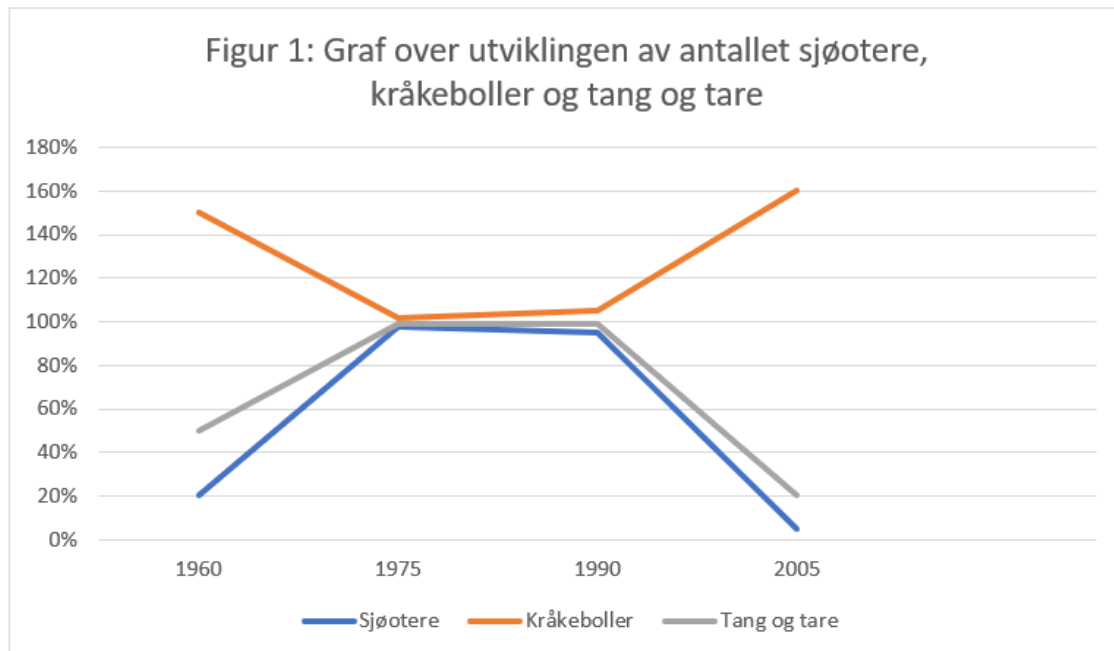
#### 3.4 a) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

 Sideskift

Side 10

#### 3.5 Undersøk grafen under, og svar på spørsmålene



Figur 1: grafen gir en oversikt over andelen av arten i prosent, i forhold til hvordan det var før mennesker påvirket økosystemet. Verdier over eller under 100% viser at det har blitt flere eller færre av arten.

### 3.5 a) Hva kan tolkes ut fra grafen? \*

- Tabellen viser at når det blir flere kråkeboller, blir det flere sjøotere og mindre tang og tare.
- Tabellen viser at nedgangen i antallet sjøotere, skjer samtidig som det blir flere kråkeboller og mindre tang og tare.
- Tabellen viser at når det blir mindre tang og tare, fører dette til at det blir det flere kråkeboller.
- Tabellen viser at når det blir færre kråkeboller, fører dette til at det blir flere sjøotere.

### 3.5 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*


Bruk informasjonen fra grafen til å svare på spørsmålet.

Hvis du var sikker på forklaringen din på oppgave 3.3 a), kan du svare det samme som du gjorde der.

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

3.5 c) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen? \*



 Sideskift

### Test av ferdigheter innen kritisk tenkning i naturfag

Denne undersøkelsen er en del av et masterprosjekt, som har som mål å utvikle en god test for kritisk tenkning i naturfag.

Ønsker du å delta i undersøkelsen? \*

- Å delta i undersøkelsen er frivillig, og du kan trekke deg om du ønsker det.
  - Svarene i undersøkelsen er anonym, og det er ikke mulig å identifisere hvem du er ut fra dine svar på testen. Svarene dine på undersøkelsen skal brukes for å videreutvikle testen, og for å undersøke hvordan ungdomsskoleelever tenker rundt kritisk tenkning.
- Jeg ønsker å delta i undersøkelsen.
- Jeg ønsker å trekke meg fra undersøkelsen, og ønsker ikke besvare testen.

Husk at du kan bla i testen, men unngå å endre på oppgaver du allerede har besvart.

På forhånd takk for at du bidrar til å utvikle testen.



Svar på spørsmålene under, før du begynner på del 1 av testen.

Tror du det kan være kreftfremkallende å spise kjøtt? \*

- Ja
- Nei

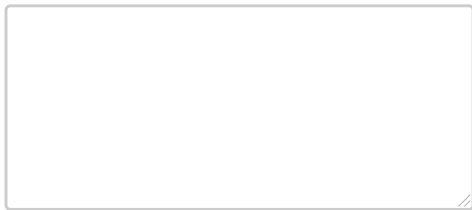
Tror du det er like kreftfremkallende å spise kjøtt som å røyke? \*

- Ja
- Nei

### Del 1 av testen: Kilder fra internett

1.0 a) Tenk at du skal være så kildekritisk som mulig når du møter en tekst fra internett. Nevn så mange ting du klarer om hva som påvirker om du stoler informasjonen på fra kilden på 2 minutter: \*





## 1.1 Kilde 1

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### "Åpne øynene

*For noen måneder siden så jeg en dokumentar som satt meg ut. "What the health" heter den, og ligger på Netflix. Den provoserte meg, men det er helt ærlig nesten da jeg liker meg selv best. Når jeg er engasjert i noe på denne måten. Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer, og jeg har SÅ lyst at dere skal se den. Men først, kan jeg komme med et par små fakta som jeg noterte meg underveis i filmen:*

*– På et slakteri slaktes 250 kuer i timen. Dette gir 15 sekunder per ku. Ting går fort, noe som har ført til at det blant annet er spør av avføring i 90% av kjøttet som selges i dagligvarebutikker i USA. (norge har – i følge et par artikler jeg fant, halvparten). For å ikke snakke om spor av puss, og all antibiotikaen og hormonene man får i seg ved å spise kjøtt / meieriprodukter. Her er ikke norge noe annerledes.*

- *Å spise kjøtt er like kreftfremkallende som å røyke.*
- *Kun 6 – 10% av alle krefttyper kan overføres genetisk. Om så, så er det ikke en eller to kreftceller som er ille, men maten du spiser kan, og VIL booste disse cellene om du spiser feil (det vil si, for mye kjøtt). Som de fleste av oss gjør..*
- *Egg er såpass lite godt for deg at det nå i amerika ikke engang er lovlig å si i reklamer at det er sunt eller bra for deg med hele setninger. Så de må heller si "a good egg", det er ikke lovlig å si "good for YOU".*
- *Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør. Dette fører til at man vil ha mer og mer, uansett om man går opp i vekt og huden tar skade.*
- *Det verste er at alt handler om kun en ting – penger. Apotekene skal tjene penger, regjeringen skal tjene penger, og kjøttindustrien skal. Disse tre jobber ofte sammen. Altså, bare SE den dokumentaren, den er så bra!*

*Det er trist at alt handler om penger. Både våre egne hoder, men samfunnet vi lever i."*



Hentet 03.09.2020 fra innlegget «Åpne øynene», Sophie Elises blogg, 10. september 2017 ([https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html))

1.1 a) Bør man stole på informasjonen i kilde 1? \*

- Ja
- Nei

1.1 b) Hvilket av alternativene om kilde 1 er du mest enig i? \*

Ordforklaring: Troverdige = Noe man kan stole på.

- Den er troverdig, fordi forfatteren skriver på en måte jeg forstår.
- Den er troverdig, fordi forfatteren har undersøkt temaet nøye, og har gode begrunnelser.
- Den er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser
- Den er ikke troverdig, fordi alt bloggere skriver er skrevet kun for å selge produkter.

### Informasjon til de neste oppgavene

Et argument består av en påstand og en begrunnelse. I eksemplene på argumenter under er påstandene markert i **gult** og begrunnelsene i **grønt**.

«**Jeg mener at det er galt å fiske med krok**, fordi **vi vet at mange dyr kan føle smerte**, og fisk er et dyr.»

«**Du burde ikke stupe fra fjellet og ut i havet**, for **det gjorde broren min, og han brakk nesa**.»

I en tekst er **påstander** budskap i teksten. **Begrunnelser** brukes for å bygge opp troverdigheten til forfatterens budskap.

1.1 c) Hvilken av påstandene under prøver kilde 1 å formidle? \*

*Ordforklaring: Formidle = Det man prøver å si*

- Å spise kjøtt er like kreftfremkallende som å røyke
- Å spise kjøtt er sunt, så lenge det man begrenser mengden.
- Å røyke er mindre kreftfremkallende enn mange tror.
- Hun har undersøkt temaet nøye, og viser til forskning.

1.1 d) Hvordan begrunner forfatteren i kilde 1 at hun funnet informasjon om temaet? \*

- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør.
- Hun har sett en dokumentar som heter "What the health" som ligger på Netflix, og søkt på google.
- Hun har undersøkt temaet ved å ta kontakt med en fagperson.
- Hun har undersøkt temaet nøye, og kan vise til forskning.

1.1 g) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.1 h) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 1 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 1.

- Enig
- Uenig

1.1 i) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 1 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

1.1 j) Er du enig i hele utsagnet under? \*

Kilde 1 er troverdig, fordi påstandene stemmer med det jeg har lært på skolen om kjøtt og røyking fra før.

- Enig
- Uenig

1.1 l) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser.

- Enig
- Uenig

1.1 m) Er du enig i hele utsagnet under? \*

Det er sant at det tar 15 sekunder å slakte en ku, hvis det slaktes 250 kuer i timen på et slakteri.

- Det er sant
- Det er ikke nødvendigvis sant



## 1.2 Kilde 2

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

**«– Ikke like kreftfremkallende**

*Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning. Direktør Giske Ursin bekrefter at det ikke er like kreftfremkallende å spise kjøtt som å røyke. Ursin tror hun vet hvor misforståelsen om at det er like kreftfremkallende å spise kjøtt som å røyke kommer fra:*



*I 2015 klassifiserte Verdens helseorganisasjon (WHO) prosessert kjøtt (som for eksempel pølser) som sikkert kreftfremkallende, mens rødt kjøtt er klassifisert som sannsynlig kreftfremkallende. Dermed kommer prosessert kjøtt nå i samme klasse som røyk. At både prosessert kjøtt og røyk er klassifisert som sikkert kreftfremkallende, er ikke det samme som at risikoen er like stor. Graden av kreftfremkallighet er ekstremt forskjellig.*

*Rødt, saltet og bearbeidet kjøtt kan øke risikoen for tarmkreft. Ifølge Kreftregisteret er det imidlertid bare en moderat andel av forekomsten av tarmkreft som kan tilskrives kosthold og livsstil. Ursin sier følgende: - For lungekreft kan derimot over 80 prosent tilskrives røyking. Direktøren understreker at Kreftregisteret støtter kreftene fra WHO som anbefaler å begrense inntaket av rødt og bearbeidet kjøtt. Dette fordi noe av kreftforekomsten kan forebygges med godt kosthold og en aktiv livsstil.»*

Hentet 03.09.2020 fra «*Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke*», Eva Akerbæk, Mina Liavik Karlsen og Silje S. Skiphamn, 15.09.2017 (<https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>)

1.2 a) Var informasjonen i kilde 2 troverdig? \*

- Ja
- Nei

1.2 b) Hva fortalte kilde 2 om kreftfaren med å spise rødt kjøtt? \*

- Det er like kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt som å røyke.
- Det er ikke kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det kan være kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det er kun bearbeidet kjøtt som er like kreftfremkallende som å røyke

1.2 c) Hvem har skrevet kilde 2? \*

- Giske Ursin
- En journalist som jobber i faktisk.no
- Det er to journalister som jobber i faktisk.no

- Det er tre journalister som jobber i faktisk.no
- En ansatt i kreftregisteret

1.2 d) Hva slags begrunnelse bruker kilde 2 for hvordan de kom frem til påstandene i teksten? \*

- Det er kun bearbeidet kjøtt som er like kreftfremkallende som å røyke
- Det begrunnes ved at faktisk.no har lest forskning på temaet.
- Faktisk.no tok kontakt med kreftregisteret, som forsker på temaet
- Det begrunnes ved at det er like kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt som å røyke.

1.2 e) Er du enig i hele utsagnet under? \*

Kilde 2 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.2 f) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 2 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 2.

- Enig
- Uenig

1.2 g) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 2 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

1.2 h) Er du enig i hele utsagnet under? \*

Påstandene i kilde 2 stemmer med det de fleste tror om kjøtt og røyking fra før, og dette styrker troverdigheten.


- Enig
- Uenig

### 1.3. Sammenligning av kilde 1 og 2

1.3 a) Hvilken kilde synes du var mest troverdig \*


- Kilde 1 (Sophie Elise «Åpne øynene»)
- Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»)

1.3 a.a Hva var de to viktigste faktorene når du synes kilde 1 var mest troverdig?

 Dette elementet vises kun dersom alternativet «Kilde 1 (Sophie Elise «Åpne øynene»» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

1.3 a.b Hva var de to viktigste faktorene når du synes kilde 2 var mest troverdig?

 Dette elementet vises kun dersom alternativet «Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

1.3 b) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det er like kreftfremkallende å spise kjøtt som å røyke»? \*

- Ja
- Nei

1.3 c) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det kan være kreftfremkallende å spise kjøtt» \*

- Ja
- Nei

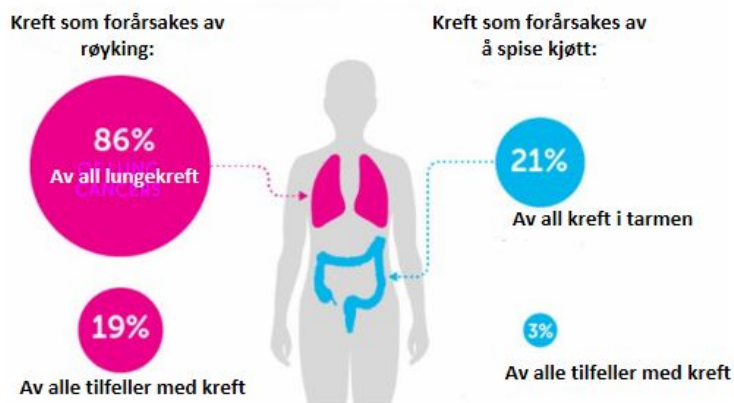
 Sideskift

Side 5

Undersøk illustrasjonen under, og svar på oppgavene

### Røyking sammenlignet med å spise kjøtt:

- Hvor stor er risikoen for å utvikle kreft?



Hvor mange får kreft av å røyke sammenlignet med å spise kjøtt:

 = 1000 tilfeller med kreft



Hvis ingen røyket ville det vært  
64 500 færre tilfeller med kreft



Hvis ingen spise kjøtt ville det vært  
8 800 færre tilfeller med kreft

Tallene er hentet fra Cancer Research UK

1.4 a) Hvor mange mennesker ville unngått å få kreft hvis ingen røykte, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500
- 21%

1.4 b) Hvor mange mennesker ville unngått å få kreft hvis ingen spiste kjøtt, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500
- 21%

1.4 c) Hvor stor andel av all lungekreft kommer av røyking, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500
- 21%

1.4 d) Hva betyr hver rosa mann i illustrasjonen? \*

- Hver rosa mann viser ett tilfelle av kreft som skyldes at man spiser kjøtt.
- Hver rosa mann viser ett tilfelle av kreft som skyldes røyking.
- Hver rosa mann viser tusen tilfeller av kreft som skyldes at de spiser kjøtt.
- Hver rosa mann viser tusen tilfeller av kreft som skyldes røyking.
- Hver rosa mann viser ett liv som ville blitt spart hvis ingen spiste kjøtt.
- Hver rosa mann viser ett liv som ville blitt spart hvis ingen røykte.



## Del 2 av testen: Vurdering av argumenter

Informasjonen under er viktig når du løser de neste 14 oppgavene.

Et **sterkt** argument kjennetegnes som:

- Direkte relevant til spørsmålet
- Bruker viktig og konkret informasjon

Et **svakt** argument kjennetegnes ved at det oppfyller et eller flere av disse kriteriene:

- Begrunnelsen eller påstanden er ikke direkte relevant for å svare på spørsmålet
- Uttrykker en mening, og ikke fakta

I oppgavene **2.1** og **2.2** skal du vurdere argumenter.

- Ta utgangspunkt i at informasjonen i teksten og argumentene er **sanne**.
- Husk at du bare skal **vurdere** argumentene, og **ikke uttrykke din mening** om saken.



## 2.1. Bør Ola kjøpe solcellepanel for å spare penger?

Ola har et stort hus, og han lurer på om han skal kjøpe solcellepanel til å ha på taket. Et solcellepanel er en energikilde som samler solenergi fra sola, og gjør den om til elektrisk energi som Ola kan bruke i huset. Da trenger ikke Ola å kjøpe strøm fra strømselskapene.

Vurder om argumentene under er svake eller sterke, og velg alternativet som passer best.

### 2.1 a) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Nei, han vil ikke spare penger på å kjøpe solcellepanel, fordi et solcellepanel er dyrt og har lav gjennomsnittlig levetid. Den går altså ofte i stykker før man har tjent inn det solcellepanelene kostet.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 b) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, det mener Per at Ola bør gjøre.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 c) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, for solcellepanel er en klimavennlig energikilde. Det er viktig at vi velger solcellepanel og andre fornybare energikilder for å stoppe global oppvarming.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 d) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, for de som har solcellepanel på taket har gjennomsnittlig bedre råd.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 e) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, banken til Ola sier at solcellepanel er en god investering

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 f) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, solcellepanel er en god investering. I løpet av fem år vil han ha spart like mye penger på strømutfgifter som han brukte på å kjøpe solcellepanelene. Etter det vil han fortsette å spare penger.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

### 2.1 g) Bør Ola kjøpe solcellepanel for å spare penger? \*

Argument:

- Ja, mange av dem som har solcellepanel er veldig fornøyde med panelene.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

## 2.2. Bør vi bruke genteknologi for å behandle sykdommer?

I genene våre står oppskriften til hvert enkelt menneske. Med fremtidens genteknologi kan vi sannsynligvis forandre oppskriften til levende organismer som bakterier, planter og dyr. Med genteknologien vil fremtidens leger kanskje kunne påvirke og forandre egenskaper hos et barn, som for eksempel øyefarge eller hvilket kjønn et barn får. Noen sykdommer blir man født med og mange av disse kan kureres ved hjelp av genteknologi.

Kan vi å bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke?

Vurder om argumentene under er svake eller sterke, og velg alternativet som passer best.

### 2.2 a) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Nei, det synes jeg ikke.

- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 b) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Ja, ved å bruke genteknologi til å kurere sykdommer vil det bli færre syke i samfunnet, og man kan bruke tiden på å behandle andre syke pasienter på sykehusene.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 c) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Ja, man bør alltid jobbe for å forbedre naturen. Samfunnet vil bli bedre, fordi genteknologi kan gjøre at mennesker blir bedre på jobbene sine.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 d) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Nei, å forandre på naturen er ikke riktig. Hvis man begynner å forandre på den genetiske oppskriften til arter, vil det bare være et spørsmål om tid før det ikke er noen grenser på hvilke endringer vi har gjort på mennesker.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 e) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Nei, den genetiske oppskriften til mennesker er komplisert, hvis vi endrer på en egenskap vil vi i mange tilfeller også endre på andre egenskaper hos mennesker, som vi ikke ønsker å endre på.
- De som blir behandlet med genteknologi kan derfor bli syke av behandlingen.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 f) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Ja, en kjent professor som heter Øystein Nilsen ser ingen utfordringer med å bruke genteknologi for å helbrede syke.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.

2.2 g) Kan vi bruke genteknologi til å behandle medfødte sykdommer i fremtiden, slik at vi får færre som er syke? \*

Argument:

- Nei, slik jeg ser det, må vi enten tillate at å forandre på alle egenskaper hos mennesker, eller så bør vi forby det.
- Sterkt argument, argumentet oppfyller kriteriene for et sterkt argument.
- Svakt argument, fordi begrunnelsen ikke er relevant for spørsmålet.
- Svakt argument, fordi begrunnelsen er en mening, og ikke fakta.



## Del 3 av testen: Å kunne trekke konklusjoner

I denne serien av oppgaver vil du stegvis bli presentert for ny informasjon om hvorfor tang og tare forsvinner fra havbunnen utenfor kysten ved Alaska.

Målet er at du ut fra denne informasjonen skal finne den mest sannsynlige forklaringen på hvorfor det skjer.

### 3.1 Bruk informasjonen, og vurder forklaringen:

- Tang og tare er sjøplanter som lever på bunnen av havet.
- På havbunnen utenfor Alaska har det gjennom flere år blitt observert at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt grunnet global oppvarming.

3.1 a) Er du enig i at konklusjonen under kan stemme? \*

**Konklusjon:** Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Ja
- Nei

3.1 b) Er det tilstrekkelig med informasjon for å trekke en konklusjon nå? \*

- Ja
- Nei

3.1 c) Hvilken annen forklaring kan du se for deg for hvorfor tang og tare forsvinner? \*

 Sideskift

Side 8

**3.2 Bruk informasjonen du har fått til å finne forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen.**

Husk at du kan bla tilbake

for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

**Ny informasjon:**

- Mengden kråkeboller har gått opp i samme tidsrom som det har blitt mindre tang og tare.
- Sjøotere er et rovdyr, og spiser kun byttedyr.
- Kråkebollen er et byttedyr.
- Tidligere var det et problem mennesker jaktet på sjøotere for å bruke pelsen deres til å lage eksklusive pelsprodukter som pelskåper. Dette førte til at sjøoterne nesten ble utryddet rundt Alaska på starten av 1900-tallet.
- Et byttedyr er maten til rovdyr. Hvis byttedyret forsvinner vil rovdyret enten dø av sult eller måtte spise en annen art.
- Spekkhoggere er en hval, og et rovdyr. Den spiser for det meste fisk. Den har ingen naturlige fiender.

3.2 a) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.2 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



**3.3 Du skal vurdere de samme forklaringene som på del 3.2, men du får mer informasjon som du kan bruke.**

Husk at du kan bla tilbake for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

**Ny informasjon:**

- Kråkeboller er planteetere.
- Et økosystem består av mange arter (altså ulike typer dyr, planter og bakterier), som påvirker hverandre.
- Det ble ulovlig å jakte på sjøotere i 1911, fordi sjøoterne var på vei til å bli utryddet.
- Mange arter er sårbare for endringer i økosystemet, som forandringer i temperatur eller hvis det kommer nye ukjente arter inn i økosystemet.

- Mennesker påvirker økosystemer blant annet gjennom global oppvarming, forsøpling av havene og ved å overfiske.
- Hvis mennesker fortsetter å kaste søppel i havene vil det i 2050 være mer plastikk i havet enn fisk.
- Tang og tare er tilhører planteriket, og er ikke et dyr.
- Sportsfiskere klager på at det har blitt vanskeligere å få fisk når de fisker.
- I 1970 var det like mange sjøotere i Alaska som det var før menneskene begynte å jakte på den. Å forby jakt på sjøotere er en suksess.

3.3 a) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

3.3 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

3.3 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

### 3.3 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

### 3.3 g) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



Side 10

### 3.4 Velg den mest sannsynlige forklaringen f or hvorfor det har blitt mindre tang og tare på havbunnen utenfor Alaska.

Husk at du kan bla tilbake

for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

#### Ny informasjon:

- En spekkhogger kan spise 1825 sjøotere i året, hvis den bare spiser otere.
- Man kan finne samme type tang som finnes på havbunnen utenfor Alaska, i områder lenger syd, hvor det er et varmere klima.
- Det har blitt 25% færre sjøotere i året på havbunnen utenfor Alaska siden 1990-tallet.
- Forskerne er forundret over at de ikke har funnet flere døde sjøotere på strender. Dette gjør at de utelukker sykdom som hovedårsaken til hvorfor det har blitt færre sjøotere, fordi syke sjøotere ville blitt skylt opp på strendene når de døde.

### 3.4 a) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

### 3.4 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.



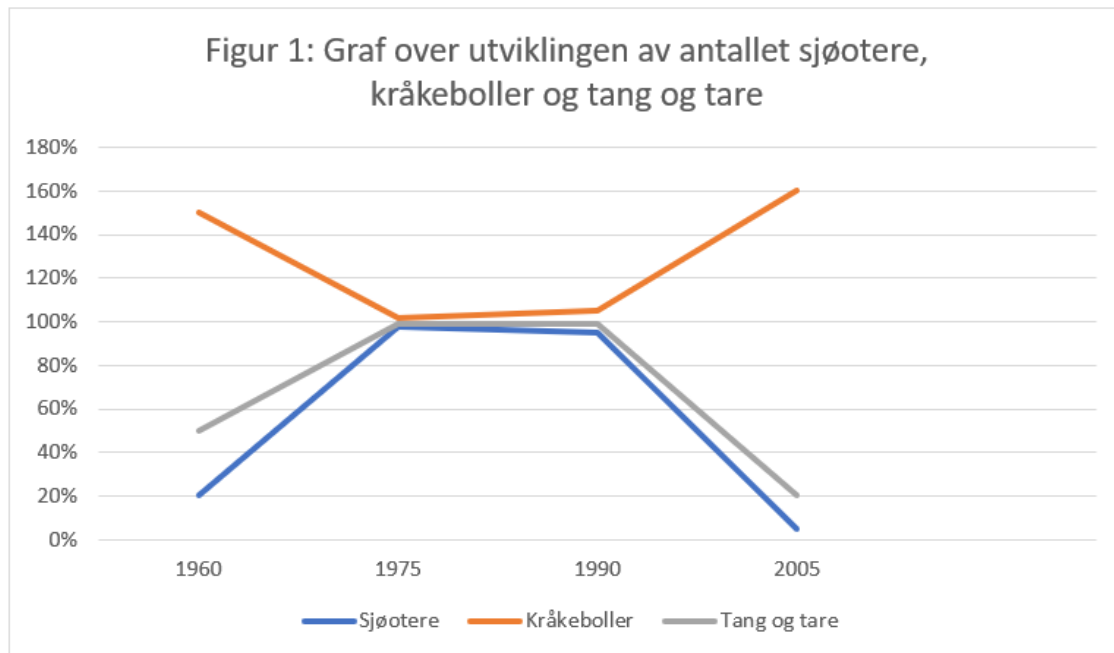
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.



Sideskift

Side 11

### 3.5 Undersøk grafen under, og svar på spørsmålene



Figur 1: grafen gir en oversikt over andelen av arten i prosent, i forhold til hvordan det var før mennesker påvirket økosystemet. Verdier over eller under 100% viser at det har blitt flere eller færre av arten.

#### 3.5 a) Hva kan tolkes ut fra grafen? \*

- Tabellen viser at når det blir flere kråkeboller, blir det flere sjøotere og mindre tang og tare.
- Tabellen viser at nedgangen i antallet sjøotere, skjer samtidig som det blir flere kråkeboller og mindre tang og tare.
- Tabellen viser at når det blir mindre tang og tare, fører dette til at det blir det flere kråkeboller.
- Tabellen viser at når det blir færre kråkeboller, fører dette til at det blir flere sjøotere.

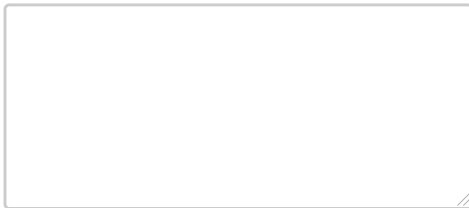
3.5 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

Bruk informasjonen fra grafen til å svare på spørsmålet.

Hvis du var sikker på forklaringen din på oppgave 3.4 b), kan du svare det samme som du gjorde der.

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

3.5 c) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen? \*



## Vedlegg til studie 3:

### Vedlegg 3.1: Kartleggingsverktøy anvendt i studie 3

Side 1

## Test av ferdigheter innen kritisk tenkning i naturfag

Denne undersøkelsen er en del av et masterprosjekt, som har som mål å utvikle en god kartleggingstest for kritisk tenkning i naturfag.

### Hva betyr det å delta i undersøkelsen?

- Å delta i undersøkelsen er frivillig, og du kan trekke deg om du ønsker det.
- Svarene i undersøkelsen er anonyme, og det er ikke mulig å identifisere hvem du er ut fra dine svar på testen. Derfor er det heller ikke mulig å trekke tilbake svarene når du har fullført hele testen. Svarene dine på undersøkelsen skal brukes for å videreutvikle testen, og for å undersøke hvordan ungdomsskoleelever tenker rundt kritisk tenkning.

Hvis du ønsker å delta i undersøkelsen kan du trykke videre i skjemaet. Ikke trykk videre om du ikke ønsker å delta.



Sideskift

Side 2

### Takk for at du ønsker å delta i undersøkelsen.

Besvar spørsmålene under:

Hvilket skoletrinn tilhører du? \*

9. trinn
10. trinn

Hvilket kjønn identifiserer du deg med? \*

- Gutt
- Jente
- Annet

Er minst en av dine foresatte født i et nordisk land (Norge, Sverige eller Danmark)? \*

- Ja
- Nei

Prøv å husk: Hvilken karakter fikk du i naturfag våren 2020? \*

- 5 eller 6
- 3 eller 4
- 1 eller 2

Prøv å husk: Hvilken karakter fikk du i norsk skriftlig (bokmål) våren 2020? \*

- 5 eller 6
- 3 eller 4
- 1 eller 2

Hvor mye leser du på fritiden? \*

- Jeg leser bare når jeg må.
- Jeg leser litt på fritiden.
- Jeg leser mye på fritiden.

Hvor utfordrende synes du det er å forstå det du leser i naturfag? \*

Ta utgangspunkt i forrige gang du **leste** i en lærebok i naturfag

- Jeg forsto lite av det jeg leste.
- Jeg forsto en del av det jeg leste.
- Jeg forsto det meste av det jeg leste.
- Jeg forsto alt av det jeg leste.



Sideskift

Side 3

Svar på spørsmålene under, før du begynner på testen.

	Ja	Nei
Tror du det kan være kreftfremkallende å spise kjøtt? *	<input type="radio"/>	<input type="radio"/>
Tror du det er like kreftfremkallende å spise kjøtt som å røyke? *	<input type="radio"/>	<input type="radio"/>



Sideskift

Side 4

## Del 1 av testen: Kilder fra internett

1.0 a) Tenk at du skal være så kildekritisk som mulig når du møter en tekst fra internett. Nevn så mange ting du klarer om hva som påvirker om du stoler informasjonen på fra kilden på 2 minutter: \*

## 1.1 Kilde 1

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### "Åpne øynene

For noen måneder siden så jeg en dokumentar som satt meg ut. "What the health" heter den, og ligger på Netflix. Den provoserte meg, men det er helt ærlig nesten da jeg liker meg selv best. Når jeg er engasjert i noe på denne måten. Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer, og jeg har SÅ lyst at dere skal se den. Men først, kan jeg komme med et par små fakta som jeg noterte meg underveis i filmen:

– På et slakteri slaktes 250 kuer i timen. Dette gir 15 sekunder per ku. Ting går fort, noe som har ført til at det blant annet er spør av avføring i 90% av kjøttet som selges i dagligvarebutikker i USA. (norge har – i følge et par artikler jeg fant, halvparten). For å ikke snakke om spor av puss, og all antibiotikaen og hormonene man får i seg ved å spise kjøtt / meieriprodukter. Her er ikke norge noe annerledes.

- Å spise kjøtt er like kreftfremkallende som å røyke.
- Kun 6 – 10% av alle krefttyper kan overføres genetisk. Om så, så er det ikke en eller to kreftceller som er ille, men maten du spiser kan, og VIL booste disse cellene om du spiser feil (det vil si, for mye kjøtt). Som de fleste av oss gjør..
- Egg er såpass lite godt for deg at det nå i amerika ikke engang er lovlig å si i reklamer at det er sunt eller bra for deg med hele setninger. Så de må heller si "a good egg", det er ikke lovlig å si "good for YOU".
- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør. Dette fører til at man vil ha mer og mer, uansett om man går opp i vekt og huden tar skade.
- Det verste er at alt handler om kun en ting – penger. Apotekene skal tjene penger, regjeringen skal tjene penger, og kjøttindustrien skal. Disse tre jobber ofte sammen. Altså, bare SE den dokumentaren, den er så bra!

Det er trist at alt handler om penger. Både våre egne hoder, men samfunnet vi lever i."



Hentet 03.09.2020 fra innlegget «Åpne øynene», Sophie Elises blogg, 10. september 2017 ([https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html))

1.1 a) Bør man stole på informasjonen i kilde 1? \*

- Ja
- Nei

1.1 b) Hvilket av alternativene om kilde 1 er du mest enig i? \*

Ordforklaring: Troverdig = Noe man kan stole på.

- Den er troverdig, fordi forfatteren skriver på en måte jeg forstår.
- Den er troverdig, fordi forfatteren har undersøkt temaet nøye, og har gode begrunnelser.

- Den er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser
- Den er ikke troverdig, fordi alt bloggere skriver er skrevet kun for å selge produkter.

### Informasjon til de neste oppgavene

Et argument består av en påstand og en begrunnelse. I eksemplene på argumenter under er påstandene markert i **gult** og begrunnelsene i **grønt**.

«**Jeg mener at det er galt å fiske med krok**, fordi **vi vet at mange dyr kan føle smerte**, og fisk er et dyr.»

«**Du burde ikke stupe fra fjellet og ut i havet**, for **det gjorde broren min, og han brakk nesa**.»

I en tekst er **påstander** budskap i teksten. **Begrunnelser** brukes for å bygge opp troverdigheten til forfatterens budskap.

1.1 c) Hvilken av påstandene under prøver kilde 1 å formidle? \*

Ordforklaring: *Formidle = Det man prøver å si*

- Å spise kjøtt er like kreftfremkallende som å røyke
- Å spise kjøtt er sunt, så lenge det man begrenser mengden.
- Å røyke er mindre kreftfremkallende enn mange tror.
- Hun har undersøkt temaet nøye, og viser til forskning.

1.1 d) Hvordan begrunner forfatteren i kilde 1 at hun funnet informasjon om temaet? \*

- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør.
- Hun har sett en dokumentar på Netflix, og søkt på google.
- Hun har undersøkt temaet ved å ta kontakt med en fagperson.
- Hun har undersøkt temaet nøye, og kan vise til forskning.

1.1 g) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.1 h) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 1 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 1.

- Enig
- Uenig

1.1 i) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 1 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

1.1 j) Er du enig i hele utsagnet under? \*

Kilde 1 er troverdig, fordi påstandene stemmer med det jeg har lært på skolen om kjøtt og røyking fra før.

- Enig
- Uenig

1.1 l) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser.

- Enig
- Uenig

1.1 m) Er utsagnet under riktig hvis det slaktes 250 kuer i timen på et slakteri? \*

Da tar det 15 sekunder for en slakter å slakte en ku, fordi  $250\text{kuer} / 60\text{minutter} = 4\text{ kuer i minuttet}$ .

- Det er sant.
- Det er ikke nødvendigvis sant.



Sideskift

## 1.2 Kilde 2

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### «– Ikke like kreftfremkallende

*Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning. Direktør Giske Ursin bekrefter at det ikke er like kreftfremkallende å spise kjøtt som å røyke. Ursin tror hun vet hvor misforståelsen om at det er like kreftfremkallende å spise kjøtt som å røyke kommer fra:*



I 2015 klassifiserte Verdens helseorganisasjon (WHO) prosessert kjøtt (som for eksempel pølser) som sikkert kreftfremkallende, mens rødt kjøtt er klassifisert som sannsynlig kreftfremkallende. Dermed kommer prosessert kjøtt nå i samme klasse som røyk. At både prosessert kjøtt og røyk er klassifisert som sikkert kreftfremkallende, er ikke det samme som at risikoen er like stor. Graden av kreftfremkallenhet er ekstremt forskjellig.

Rødt, saltet og bearbeidet kjøtt kan øke risikoen for tarmkreft. Ifølge Kreftregisteret er det imidlertid bare en moderat andel av forekomsten av tarmkreft som kan tilskrives kosthold og livsstil. Ursin sier følgende: - For lungekreft kan derimot over 80 prosent tilskrives røyking. Direktøren understreker at Kreftregisteret støtter kreftrådene fra WHO som anbefaler å begrense inntaket av rødt og bearbeidet kjøtt. Dette fordi noe av kreftforekomsten kan forebygges med godt kosthold og en aktiv livsstil.»

Hentet 03.09.2020 fra «Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke», Eva Akerbæk, Mina Liavik Karlsen og Silje S. Skiphamn, 15.09.2017 (<https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>)

1.2 a) Var informasjonen i kilde 2 troverdig? \*

- Ja
- Nei

1.2 b) Hva fortalte kilde 2 om kreftfaren med å spise rødt kjøtt? \*

- Det er like kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt som å røyke.
- Det er ikke kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det kan være kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det er kun bearbeidet kjøtt som er like kreftfremkallende som å røyke

1.2 c) Hvem har skrevet kilde 2? \*

- Giske Ursin
- En journalist som jobber i faktisk.no
- Det er to journalister som jobber i faktisk.no
- Det er tre journalister som jobber i faktisk.no
- En ansatt i kreftregisteret
- To ansatte i kreftregisteret
- Tre ansatte i kreftregisteret

1.2 d) Hva slags begrunnelse bruker kilde 2 for hvordan de kom frem til påstandene i teksten? \*



- Det er kun bearbejdet kjøtt som er like kreftfremkallende som å røyke
- Det begrunnes ved at faktisk.no har lest forskning på temaet.
- Faktisk.no tok kontakt med kreftregisteret, som forsker på temaet
- Det begrunnes ved at det er like kreftfremkallende å spise rødt, saltet og bearbejdet kjøtt som å røyke.

1.2 e) Er du enig i hele utsagnet under? \*

Kilde 2 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.2 f) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 2 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 2.

- Enig
- Uenig

1.2 g) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 2 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

1.2 h) Er du enig i hele utsagnet under? \*

Påstandene i kilde 2 stemmer med det de fleste tror om kjøtt og røyking fra før, og dette styrker troverdigheten.

- Enig
- Uenig



Sideskift

### 1.3. Sammenligning av kilde 1 og 2

**Tips:**

- Husk at du kan bla tilbake i testen, men unngå å endre på oppgaver du allerede har besvart.

1.3 a) Hvilken kilde synes du var mest troverdig \*

- Kilde 1 (Sophie Elise «Åpne øynene»)
- Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»)

### 1.3 a.a Hva var de to viktigste faktorene når du synes kilde 1 var mest troverdig?

Dette elementet vises kun dersom alternativet «Kilde 1 (Sophie Elise «Åpne øynene»）」 er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

### 1.3 a.b Hva var de to viktigste faktorene når du synes kilde 2 var mest troverdig?

Dette elementet vises kun dersom alternativet «Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»）」 er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

### 1.3 b) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det er like kreftfremkallende å spise kjøtt som å røyke»? \*

- Ja
- Nei

### 1.3 c) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det kan være kreftfremkallende å spise kjøtt»? \*

- Ja
- Nei

1.3 d) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

Å spise kjøtt er ikke bra for klimaet, og derfor bør man unngå det.

- Argumentet er relevant
- Argumentet er ikke relevant

1.3 e) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

Det er mange næringsstoffer i kjøtt som mennesker trenger, og derfor er det sunt.

- Argumentet er relevant
- Argumentet er ikke relevant



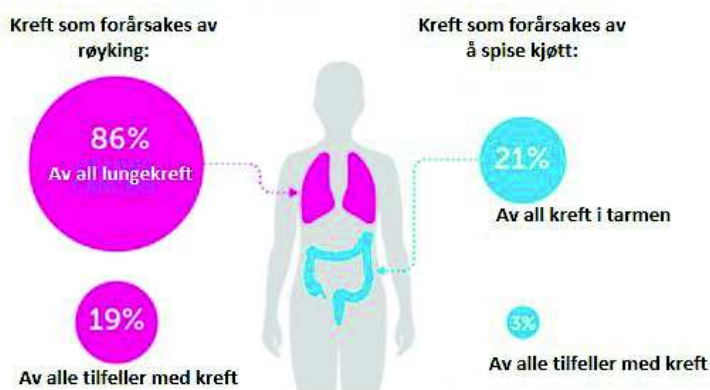
Sideskift

Side 7

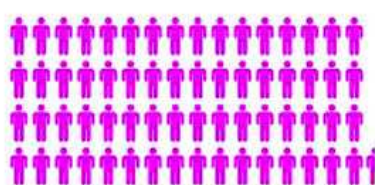
Undersøk illustrasjonen under, og svar på oppgavene

### Røyking sammenlignet med å spise kjøtt:

- Hvor stor er risikoen for å utvikle kreft?



Hvor mange får kreft av å røyke sammenlignet med å spise kjøtt:



Hvis ingen røyket ville det vært 64 500 færre tilfeller med kreft

= 1000 tilfeller med kreft



Hvis ingen spiste kjøtt ville det vært 8 800 færre tilfeller med kreft

Tallene og illustrasjonen er hentet fra Cancer Research UK

1.4 a) Hvor mange mennesker ville unngått å få kreft hvis ingen røykte, ifølge illustrasjonen? \*

- 86%
- 8 800

64 500

21%

1.4 b) Hvor mange mennesker ville unngått å få kreft hvis ingen spiste kjøtt, ifølge illustrasjonen? \*

86%

8 800

64 500

21%

1.4 c) Hvor stor andel av all lungekreft kommer av røyking, ifølge illustrasjonen? \*

86%

8 800

64 500

21%

3%

1.4 d) Hvor stor andel av alle krefttilfeller skyldes av at folk spiser kjøtt, ifølge illustrasjonen? \*

21%

19 %

64 500

86 %

3 %

1.4 e) Hva betyr hver rosa mann i illustrasjonen? \*

Hver rosa mann viser ett tilfelle av kreft som skyldes at man spiser kjøtt.

Hver rosa mann viser ett tilfelle av kreft som skyldes røyking.

Hver rosa mann viser tusen tilfeller av kreft som skyldes at de spiser kjøtt.

Hver rosa mann viser tusen tilfeller av kreft som skyldes røyking.

Hver rosa mann viser ett liv som ville blitt spart hvis ingen spiste kjøtt.

Hver rosa mann viser ett liv som ville blitt spart hvis ingen røykte.

## Del 2 av testen: Å kunne trekke konklusjoner

I denne serien av oppgaver vil du stegvis bli presentert for ny informasjon om hvorfor tang og tare forsvinner fra havbunnen utenfor kysten ved Alaska.

Målet er at du ut fra denne informasjonen skal finne den mest sannsynlige forklaringen på hvorfor det skjer.

### 2.1 Bruk informasjonen, og vurder forklaringen:

- Tang og tare er sjøplanter som lever på bunnen av havet.
- På havbunnen utenfor Alaska har det gjennom flere år blitt observert at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt grunnet global oppvarming.

2.1 a) Er du enig i at konklusjonen under kan stemme? \*

**Konklusjon:** Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Ja
- Nei

2.1 b) Er det tilstrekkelig med informasjon for å trekke en konklusjon nå? \*

- Ja
- Nei

2.1 c) Hvilken annen forklaring kan du se for deg for hvorfor tang og tare forsvinner? \*



Sideskift

## 2.2 Bruk informasjonen du har fått til å finne forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen.

Husk at du kan bla tilbake

for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

### Ny informasjon:

- Antallet kråkeboller har gått opp i samme tidsrom som det har blitt mindre tang og tare.
- Sjøotere er et rovdyr, og spiser kun byttedyr.
- Kråkebollen er et byttedyr.
- Tidligere var det et problem mennesker jaktet på sjøotere for å bruke pelsen deres til å lage eksklusive pelsprodukter som pelskåper. Dette førte til at sjøoterne nesten ble utryddet rundt Alaska på starten av 1900-tallet.
- Et byttedyr er maten til rovdyr. Hvis byttedyret forsvinner vil rovdyret enten dø av sult eller måtte spise en annen art.

- Spekkhoggere er en hval, og et rovdyr. Den spiser for det meste fisk. Den har ingen naturlige fiender.

2.2 a) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.2 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.2 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.2 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.2 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.2 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



### 2.3 Du skal vurdere de samme forklaringene som på del 3.2, men du får mer informasjon som du kan bruke.

Husk at du kan bla tilbake

for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

#### Ny informasjon:

- Kråkeboller er planteetere.
- Et økosystem består av mange arter (altså ulike typer dyr, planter og bakterier), som påvirker hverandre.
- Det ble ulovlig å jakte på sjøotere i 1911, fordi sjøotene var på vei til å bli utryddet.
- Mange arter er sårbare for endringer i økosystemet, som forandringer i temperatur eller hvis det kommer nye ukjente arter inn i økosystemet.
- Mennesker påvirker økosystemer blant annet gjennom global oppvarming, forsøpling av havene og ved å overfiske.
- Hvis mennesker fortsetter å kaste søppel i havene vil det i 2050 være mer plastikk i havet enn fisk.
- Tang og tare er tilhører planteriket, og er ikke et dyr.
- Sportsfiskere klager på at det har blitt vanskeligere å få fisk når de fisker.
- I 1970 var det like mange sjøotere i Alaska som det var før menneskene begynte å jakte på den. Å forby jakt på sjøotere er en suksess.

2.3 a) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

2.3 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen. \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.

- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 g) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



Sideskift

Side 11

**2.4 Velg den mest sannsynlige forklaringen f  
or hvorfor det har blitt mindre tang og tare på havbunnen utenfor Alaska.**

Husk at du kan bla tilbake for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

**Ny informasjon:**

- En spekkhogger kan spise 1825 sjøotere i året, hvis den bare spiser otere.
- Man kan finne samme type tang som finnes på havbunnen utenfor Alaska, i områder lenger syd, hvor det er et varmere klima.
- Det har blitt 25% færre sjøotere i året på havbunnen utenfor Alaska siden 1990-tallet.
- Forskerne er forundret over at de ikke har funnet flere døde sjøotere på strender. Dette gjør at de utelukker sykdom som hovedårsaken til hvorfor det har blitt færre sjøotere, fordi syke sjøotere ville blitt skylt opp på strendene når de døde.

2.4 a) Har du nok informasjon til å utelukke de fleste forklaringene om hvorfor det har blitt mindre tang og tare? \*



- Ja
- Nei

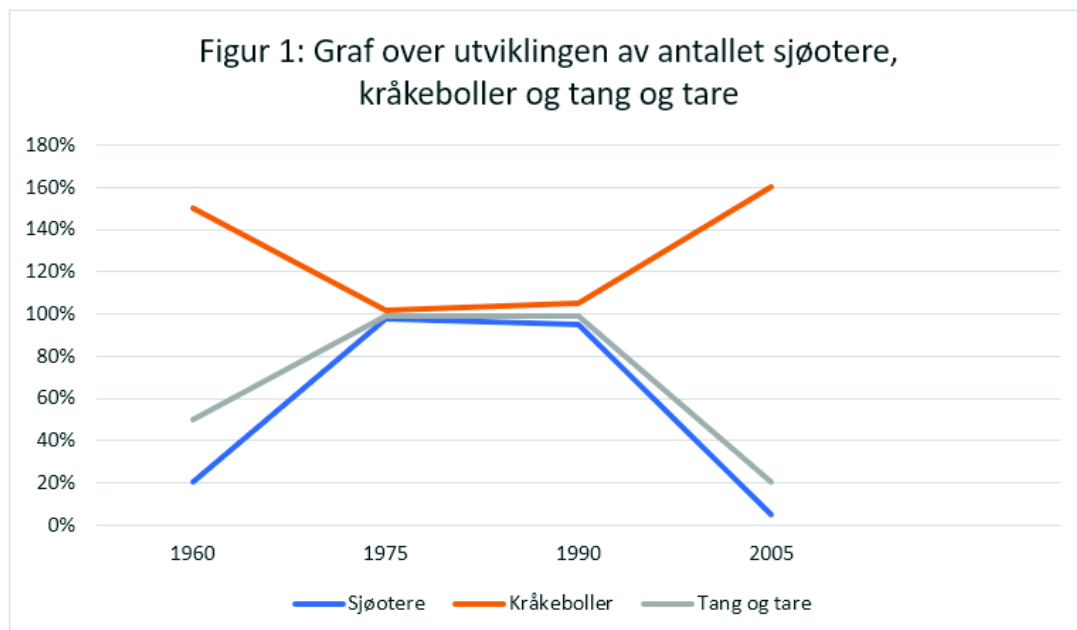
2.4 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

 Sideskift

Side 12

2.5 Undersøk grafen under, og svar på spørsmålene



Figur 1: grafen gir en oversikt over andelen av arten i prosent, i forhold til hvordan det var før mennesker påvirket økosystemet. Verdier over eller under 100% viser at det har blitt flere eller færre av arten.

2.5 a) Hva er viser figur 1? \*

- Figur 1 viser en oversikt over temperaturer i havområdet utenfor Alaska fra 1960 til 2005.
- Figur 1 viser en oversikt over temperaturer i havområdet utenfor Alaska fra 1960 til i dag.
- Figur 1 viser en graf over utviklingen av antallet sjøotere, kråkeboller og tang og tare utenfor Alaska, i tidsperioden 1960 til i dag.
- Figur 1 viser en graf over utviklingen av antallet sjøotere, kråkeboller og tang og tare utenfor Alaska, i tidsperioden 1960 til 2005.

2.5 a) Hva kan tolkes ut fra grafen i figur 1? \*

- Tabellen viser at når det blir flere kråkeboller, blir det flere sjøotere og mindre tang og tare.
- Tabellen viser at nedgangen i antallet sjøotere, skjer samtidig som det blir flere kråkeboller og mindre tang og tare.
- Tabellen viser at når det blir mindre tang og tare, fører dette til at det blir det flere kråkeboller.
- Tabellen viser at når det blir færre kråkeboller, fører dette til at det blir flere sjøotere.

2.5 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

Bruk informasjonen fra grafen til å svare på spørsmålet.

Hvis du var sikker på forklaringen din på oppgave 3.4 b), kan du svare det samme som du gjorde der.

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

2.5 c) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen? \*

## Hvor enig er du i påstandene?

Spørsmålene har ikke et fasitsvar. Svar ut fra hvor mye du kjenner deg igjen i påstandene.

	Helt enig	Litt enig	Verken enig eller uenig	Litt uenig	Helt uenig
Å studere nye ting hele livet ville vært fantastisk . *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Andre har rett på sine meninger, men jeg trenger ikke høre dem . *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er stolt av hvor flink jeg er til å komme opp med kreative løsninger på problemer. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det viktigste i en diskusjon er å vinne. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Når jeg skal ta en avgjørelse, forsøker jeg først å finne mest mulig informasjon om saken. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er nysgjerrig på det vi lærer i naturfag. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å lære å tenke kritisk er det noe av det viktigste man lærer på skolen. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ofta legger jeg lite innsats i å lære noe nytt på skolen, før jeg vurderes i det. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg kan gjøre narr av og erte andre for deres meninger og verdier. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Når en oppgave er komplisert får jeg en følelse av panikk. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uansett hva som skjer bør man holde på verdiene sine, og ikke revurdere dem. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I stedet for å jobbe grundig med å	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

løse en oppgave, spør jeg en klassekamerat om svaret. \*



Jeg synes det man lærer i naturfag er viktig å kunne når man skal delta i demokratiske valg. \*



Jeg har lært mye hvordan man skal tenke kritisk på skolen. \*



Jeg bruker mye tid på å lese om ting jeg er nysgjerrig på. \*



Jeg er åpen for ny informasjon, selv om jeg er ganske sikker i saken. \*



Jeg gir meg ikke når oppgavene er vanskelige. \*



Det er viktig for meg at den kunnskapen jeg har er så sikker som mulig. \*



Jeg tar beslutninger raskt, og tenker på konsekvensene først når de kommer. \*



På fritiden tenker jeg over ting jeg lærer i naturfag. \*



På skolen har jeg lært hvordan man kan være kildekritisk. \*



Jeg kan bruke timevis på å lese om ukjente temaer på internett. \*



Jeg liker å lese på internett om meninger jeg er uenig med, for å få innsikt i hvordan andre tenker. \*



Jeg er god til å vurdere om man bør stole på en nettside eller ikke. \*



Jeg velger ut informasjon som passer med det jeg mener. \*



Jeg sjekker alltid med flere kilder når jeg leser om et tema jeg kan lite om fra før. \*



Naturfagskunnskapene jeg har hjelper meg med å være kildekritisk \*



### Vedlegg 3.2: Resultater på reliabilitetsanalyse av holdningskategoriene i studie 3

Holdningskategori	Cronbach alfa	Variabelnavn	Cronbach alfa Hvis slettet	Tilknyttede spørsmål	Mulig feil	Forslag til endring
<b>Intellektuell nysgjerrighet</b> (Inquisitiveness)	0,26	I-1	<b>0,38</b>	Å studere nye ting <b>hele livet ville vært fantastisk *</b> .	Flerdimensjonal	Å lære om nye temaer er spennende.
		I-2	<b>0,33</b>	Ofte legger jeg lite innsats i å lære noe nytt på skolen, <b>før jeg vurderes i det.</b> (negativ)**	Flerdimensjonal	Ofte legger jeg lite innsats i å lære noe nytt på skolen. (negativ)
		I-3	-0,42	Jeg bruker mye tid på å lese om ting jeg er nysgjerrig på.		
		I-4	0,29	Jeg kan bruke timevis på å lese om ukjente temaer <b>på internett.</b> ***	Flerdimensjonal	Jeg liker å lese om temaer jeg vet lite om.
<b>Åpenhet</b> for ny informasjon (Open-Mindedness)	0,15	A-1	<b>0,44</b>	Andre har rett på sine meninger, <b>men jeg trenger ikke høre dem.</b> (negativ) *	Flerdimensjonal	Alle har rett på sin mening.
		A-2	<b>0,32</b>	Jeg kan gjøre <b>narr av og erte</b> andre for deres meninger og verdier. (negativ) *	Relevans	Jeg viser respekt for andres meninger.
		A-3	0,07	Jeg er åpen for ny informasjon, selv om jeg er ganske sikker i saken.		
		A-4	-0,77	Jeg liker å lese på internett om meninger jeg er uenig med, for å få innsikt i hvordan andre tenker. ***		
<b>Selvtilit</b> (Critical Thinking Self-Confidence og Analyticity)	-0,03	Se-1	-0,41	Jeg er <b>stolt av</b> hvor flink jeg er til å komme opp med kreative løsninger på problemer. *	Flerdimensjonal	Jeg er flink til å finne kreative løsninger på oppgaver.

		Se-2	-0,03	Når en oppgave er komplisert, får jeg en følelse av panikk. (negativ) *	Klarhet	Jeg får en følelse av panikk når oppgaver er vanskelige <b>(negativ)</b>
		Se-3	<b>0,30</b>	Jeg gir meg <b>ikke</b> når oppgavene er vanskelige.	Negativt ord	Jeg gir meg når oppgavene er vanskelige <b>(negativ)</b>
		Se-4	0,03	Jeg er god til å vurdere om man bør stole på en nettside eller <b>ikke</b> . ***	Negativt ord / Klarhet / Holdningsobjekt for spesifikt	Jeg reflekterer godt.
<b>Sannhetssøkende</b> (Truth-Seeking)	0,51	Sa-1	0,36	Det viktigste i en diskusjon er å vinne. (negativ) **		
		Sa-2	0,36	Uansett hva som skjer bør man holde på verdiene sine, og ikke revurdere dem. (negativ)		
		Sa-3	<b>0,59</b>	Det er viktig for meg at den kunnskapen jeg har er så sikker som mulig.		
		Sa-4	0,33	Jeg velger ut informasjon som passer med det jeg mener. (negativ) ** (***)		
<b>Betenksomhet</b> ovenfor avgjørelser  (Systematicity og Maturity)	0,36	B-1	<b>0,42</b>	Når jeg skal ta en avgjørelse, forsøker jeg først å <b>finne mest mulig</b> informasjon om saken. *	Flerdimensjonal, Klarhet	Jeg samler mye informasjon før jeg tar en avgjørelse
		B-2	-0,06	I stedet for å jobbe grundig med å løse en oppgave, spør jeg en klassekamerat om svaret. (negativ) **		
		B-3	<b>0,43</b>	Jeg <b>tar beslutninger raskt</b> , og tenker på konsekvensene først når de kommer. <b>(negativ)</b>	Flerdimensjonal	Jeg vurderer konsekvensene før jeg tar en avgjørelse.
		B-4	0,31	Jeg sjekker <b>alltid</b> med flere kilder, når jeg leser om et tema jeg kan lite om fra før. ***	Flerdimensjonal Bruk av ordet «Alltid»	Jeg bruker flere kilder når jeg undersøker et tema
Holdninger til <b>naturfag</b>	0,81	N-1	0,64	Jeg er nysgjerrig på det vi lærer i naturfag.		

		N-2	0,78	Jeg synes det man lærer i naturfag er viktig å kunne når man skal delta i demokratiske valg.		
		N-3	<b>0,84</b>	På fritiden tenker jeg over ting jeg lærer i naturfag.		
		N-4	0,75	Naturfagskunnskapene jeg har hjelper meg med å være kildekritisk. ***		
Holdninger til viktigheten av <b>kritisk tenkning</b> i skolen	0,59	H-1	0,39	Å lære å tenke kritisk er det noe av det viktigste man lærer på skolen.		
		H-2	0,30	Jeg har lært mye hvordan man skal tenke kritisk på skolen.		
		H-3	<b>0,68</b>	På skolen har jeg lært hvordan man kan være kildekritisk. ***		

## Vedlegg til studie 4:

### Vedlegg 4.1: Oversikt over kategoriene, delferdighetene og oppgavene i studie 4

Kategori i testen	Beskrivelse	Operasjonalisering / Hvordan måles dette?	Delferdigheter	Oppgaveform	Tilknyttede oppgaver	n oppgaver (oppgavetype)
<b>Strategier for kritisk tenkning (S)</b>	Evnen til å ha og bruke strategier knyttet til å kunne ta en avgjørelse og utøve kildekritikk. Denne ferdigheten inkluderer evnen til å kunne gjøre rede for valg man har tatt og strategier man har brukt.  Kategorien har tatt utgangspunkt i ferdighetene «Forklare/explantation» og «Selvregulering/self-regulation», og delferdigheten «vurdering av troverdighet» (Facione 1990).	I testen skal eleven gjøre rede for hvordan man bør tenke når man skal ta en avgjørelse eller vurdere en kildes troverdighet.  Oppgavene tester om eleven gjøre rede for strategier og om eleven klarer å vurdere en kilde ut fra kriterier.	<b>S.1 Strategier</b> <ul style="list-style-type: none"> <li>• Kan begrunne og reflektere over egne strategier for hvordan man selv tar en gjennomtenkt konklusjon.</li> <li>• Kan forklare hvilke strategier man bruker for å være kildekritisk.</li> </ul>	Nevne faktorer om kildekritikk (åpen):	1.0 a)	2 (åpne)
				Forklarer hvordan man tenkte for å konkludere hva som var den mest sannsynlige forklaringen (åpen)	2.4 c)	
			<b>S.2 Vurdere troverdighet</b> <ul style="list-style-type: none"> <li>• Kan vurdere en kildes troverdighet gjennom å stille enkle spørsmål om en tekst.</li> </ul>	Enig/uenig: Vurdering om kilden er troverdig (2m)	1.1 a) 1.2 a) 1.3 a)	12 (2m) 1 (4m)
				Vurdere fire påstander om kildens troverdighet (4m)	1.1 b)	
<b>Forståelse for</b>	Evnen til å kunne forstå naturfaglig informasjon i	I testen vil eleven bli presentert for informasjon	<b>N.1</b>	Kjenne igjen påstand og begrunnelse (4m):	1.1 c), 1.1 d), 1.2 b), 1.2 d)	5 (4m)



<b>naturfaglig informasjon (N)</b>	ulike former, og å kunne forstå hvilken rolle informasjon er ment å ha i en tekst. Dette krever at eleven avkoder informasjonen, og forstår hvilken kontekst informasjonen er uttrykt.  Eleven må også kunne identifisere argumenter i tekster, og forstå hvilken konklusjon de er ment å støtte.  Kategorien har tatt utgangspunkt i ferdighetene «tolke/interpretation» og «analysere/analysis» (Facione 1990).	i form av utsnitt fra to nettbaserte tekster, tabeller og informasjon om et økosystem tilknyttet en case som elevene skal løse.  Oppgavene tester elevens forståelse av denne informasjonen.	<b>Avkode og tydeliggjøre informasjon fra tekster</b> <ul style="list-style-type: none"> <li>• Kan vise forståelse for naturfagsrelevant informasjon fra nettbaserte tekster.</li> <li>• Kan identifisere argumenter, påstander og begrunnelser i nettbaserte tekster.</li> </ul>	Spørsmål om hvem som har skrevet artiklene (4m)	1.2c)	
			<b>N.2 Avkode informasjon fra tabeller og illustrasjoner</b> <ul style="list-style-type: none"> <li>• Kan lese, avkode og analysere informasjon ut fra tabeller og illustrasjoner.</li> </ul>	Tolke og forstå informasjon fra illustrasjon om kreftfaren ved kjøttforbruk og røyking (4m)	1.4 a), 1.4 b), 1.4 c), 1.4 d), 1.4 e)	7 (4m)
	Tolke og forstå graf om utvikling av antall arter (4m)	2.5 a) 2.5 b)				
<b>Trekke velbegrunnede konklusjoner (K)</b>	Eleven identifisere hva som trengs for å trekke troverdige konklusjoner, og bruke relevant informasjon for å trekke gjennomtenkte slutninger. Dette krever at man forstår de logiske konsekvensene som kan utledes fra informasjon.  Kategorien har tatt utgangspunkt i ferdighetene	I testen blir elevene presentert for en case, og blir gradvis introdusert for informasjon om denne casen.  Oppgavene tester om elevene klarer å systematisk vurdere flere mulige forklaringer på casen, med grunnlag i	<b>K.1 Undersøke og vurdere bevis</b> <ul style="list-style-type: none"> <li>• Kan vurdere hvilke forklaringer som er mulige med grunnlag i bevis.</li> </ul>	Enig/uenig med vurdering av mulig forklaring (2m)	2.1 a) 2.2 a) - f) 2.3 b) - g)	13 (2m)
			<b>K.2 Vurdere bevisgrunnlag</b> <ul style="list-style-type: none"> <li>• Kan formulere egne alternative forklaringer til problemstillinger.</li> </ul>	Kan formulere egne alternative forklaringer (åpen).	2.1 c)	1 åpen 3 (2m)

	«Å trekke konklusjoner/inference» og «vurdere/evaluating» (Facione 1990).	bevis. I tillegg skal elevene klare å formulere alternative forklaringer på casen, og vurdere når de har nok bevisgrunnlag for å fatte en avgjørelse. To oppgaver om kjøtt og kreft er også knyttet til ferdigheten.	<ul style="list-style-type: none"> <li>Kan vurdere om det det er tilstrekkelig med informasjon for å ta en avgjørelse.</li> </ul>	Kan ta avgjøre om det er tilstrekkelig med informasjonsgrunnlag for å ta et var (2m)	2.1 b) 2.3 a) 2.4 a)	
			<b>K.3 Konkludere</b> <ul style="list-style-type: none"> <li>Kan trekke nøye gjennomtenkte konklusjoner, med grunnlag i bevis.</li> </ul>	Enig/uenig: Kunne konkludere om påstander om kjøtt og kreftfare (2m)	1.3 b) 1.3 c)	2 (2m) 2 (6m)
				Avgjøre den mest sannsynlige forklaringen (6m)	2.4 b) 2.5 b)	
			<b>K 4 Vurdere argumenter</b> <ul style="list-style-type: none"> <li>Med utgangspunkt i relevans og logisk holdbarhet</li> </ul>	Vurdere relevans eller logisk styrke	1.1 J), 1.3 d) - g)	5 (2m)

(m) = mulige svar/ alternativer

## Vedlegg 4.2: Løsningsforslag og poengutdeling på testens åpne oppgaver

Formulering åpne oppgaver	Max score på oppgaven	Kriterier åpne oppgaver
<p><b>1.0 a)</b> Tenk at du skal være så kildekritisk som mulig når du møter en tekst fra internett. Nevn så mange ting du klarer om hva som påvirker om du stoler informasjonen på fra kilden på 2 minutter: *</p>	3	<p><b>0. Poeng:</b> Eleven uttrykker ingen strategier for hvordan han kan undersøke en kildes troverdighet. Typisk svar er «<i>vet ikke</i>»</p> <p><b>1. Poeng:</b> Eleven uttrykker en eller flere strategier for å undersøke en kildes troverdighet, men faktorene eleven nevner samsvarer ikke med forslagene under på faktorene som påvirker en kildes troverdighet.</p> <p><b>2. Poeng:</b> Eleven uttrykker flere strategier for å undersøke en kildes troverdighet, og minst 2 av disse samsvarer med forslagene under på faktorene som påvirker en kildes troverdighet. Det eleven skriver er så lite spesifikt og upresist at det kan være feil.</p> <p><b>3. Poeng:</b> Eleven uttrykker flere strategier for å undersøke en kildes troverdighet, og minst 2 av disse samsvarer med forslagene under på faktorene som påvirker en kildes troverdighet. Hvis to av strategiene samsvarer må disse være presist formulert, og hvis eleven uttrykker flere strategier som samsvarer utdeles 3 poeng.</p> <p><b>Faktorer:</b> (Basert på Ennis (1985), Walton (2005) og Weyergang og Frønes (2020))</p> <ul style="list-style-type: none"> <li>- Ekspertise</li> <li>- Fagperson/autoritet på emnet</li> <li>- Bruk av forskning (jo mer om den vitenskapelige prosessen jo bedre)</li> <li>- Ingen interessekonflikter</li> <li>- God evne til å forklare hvordan man kom frem til en konklusjon.</li> <li>- Hensikten med nettsiden (informere, underholde, selge?)</li> <li>- Forsiktige konklusjoner og påstander (i motsetning til vågale) <ul style="list-style-type: none"> <li>○ Bryter det med det de fleste tror er sant?</li> </ul> </li> <li>- Omdømme til nettsiden (Hvorfor er FHI for eksempel troverdig)</li> </ul>

		<ul style="list-style-type: none"> <li>- Sammenligne med andre nettsider/eksperter.</li> <li>- Andre faktorer: Når er nettsiden fra, hvor troverdig ser den ut</li> </ul>
2.1 c) Hvilken annen forklaring kan du se for deg for hvorfor tang og tare forsvinner?	2	<p><b>0. Poeng:</b> Eleven uttrykker ikke en alternativ forklaring. Typisk svar er «vet ikke» eller om eleven gjentar forklaringen fra 2.1 a).</p> <p><b>1. Poeng:</b> Eleven uttrykker en alternativ forklaring. Forklaringen viser ikke til en sammenheng til hvorfor det blir mindre tang og tare.</p> <p><b>2. Poeng:</b> Eleven uttrykker en alternativ forklaring. Forklaringen viser til en sammenheng til hvorfor det blir mindre tang og tare.</p>
2.4 c) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen?	3	<p><b>0. Poeng:</b> Eleven uttrykker ingen strategier for hvordan han kom frem til den riktige forklaringen. Typisk svar er «vet ikke»</p> <p><b>1. Poeng:</b> Eleven uttrykker en begrunnelse for forklaringen, men uttrykker ingen konkrete strategier.</p> <p><b>2. Poeng:</b> Eleven uttrykker en forklaring for hvordan han tenkte gjennom oppgaven, men uttrykker ikke noen systematisk strategi for å finne den mest sannsynlige forklaringen gjennom del 2.</p> <p><b>3. Poeng:</b> Eleven uttrykker en forklaring for hvordan han tenkte gjennom oppgaven, og uttrykker noen systematisk strategi for å finne den mest sannsynlige forklaringen gjennom del 2.</p>

### Test av ferdigheter innen kritisk tenkning i naturfag

Denne undersøkelsen er en del av et masterprosjekt, som har som mål å utvikle et godt kartleggingsverktøy for kritisk tenkning i naturfag.

#### Hva betyr det å delta i undersøkelsen?

- Å delta i undersøkelsen er frivillig, og du kan trekke deg om du ønsker det.
- Svarene i undersøkelsen er anonyme, og det er ikke mulig å identifisere hvem du er ut fra dine svar på testen. Derfor er det heller ikke mulig å trekke tilbake svarene når du har fullført hele testen. Svarene dine på undersøkelsen skal brukes for å videreutvikle testen, og for å undersøke hvordan ungdomsskoleelever tenker rundt kritisk tenkning.

Hvis du ønsker å delta i undersøkelsen kan du trykke videre i skjemaet. Ikke trykk videre om du ikke ønsker å delta.



Sideskift

#### Takk for at du ønsker å delta i undersøkelsen.

Besvar spørsmålene under:

Hvilket skoletrinn tilhører du? \*

9. trinn
10. trinn

Hvilket kjønn identifiserer du deg med? \*

- Gutt
- Jente
- Annet

Er minst en av dine foresatte født i et nordisk land (Norge, Sverige eller Danmark)? \*

- Ja
- Nei

Prøv å husk: Hvilken karakter fikk du i naturfag høsten 2020? \*

- 5 eller 6
- 3 eller 4
- 1 eller 2

Prøv å husk: Hvilken karakter fikk du i norsk skriftlig (bokmål) høsten 2020? \*

- 5 eller 6
- 3 eller 4
- 1 eller 2

Hvor mye leser du på fritiden? \*

- Jeg leser bare når jeg må.
- Jeg leser litt på fritiden.
- Jeg leser mye på fritiden.

Ta utgangspunkt i forrige gang du leste i en lærebok i naturfag. Hvor mye forsto du av det du leste? \*

- Jeg forsto lite av det jeg leste.
- Jeg forsto en del av det jeg leste.
- Jeg forsto det meste av det jeg leste.
- Jeg forsto alt av det jeg leste.



Side 3

Svar på spørsmålene under, før du begynner på testen.

	Ja	Nei
Tror du det kan være kreftfremkallende å spise kjøtt? *	<input type="radio"/>	<input type="radio"/>
Tror du det er like kreftfremkallende å spise kjøtt som å røyke? *	<input type="radio"/>	<input type="radio"/>



Side 4

## Del 1 av testen: Kilder fra internett

1.0 a) Tenk at du skal være så kildekritisk som mulig når du møter en tekst fra internett: Skriv ned så mange ting som mulig som påvirker om du stoler på kilden. \*

Bruk maksimalt 2 minutter på denne oppgaven. Svar gjerne punktvis.

## 1.1 Kilde 1

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### "Åpne øynene

For noen måneder siden så jeg en dokumentar som satt meg ut. "What the health" heter den, og ligger på Netflix. Den provoserte meg, men det er helt ærlig nesten da jeg liker meg selv best. Når jeg er engasjert i noe på denne måten. Etter å ha brukt en drøy time på dokumentaren har jeg googlet rundt i flere timer, og jeg har SÅ lyst at dere skal se den. Men først, kan jeg komme med et par små fakta som jeg noterte meg underveis i filmen:

- På et slakteri slaktes 250 kuer i timen. Dette gir 15 sekunder per ku. Ting går fort, noe som har ført til at det blant annet er spor av avføring i 90% av kjøttet som selges i dagligvarebutikker i USA. (norge har – i følge et par artikler jeg fant, halvparten). For å ikke snakke om spor av puss, og all antibiotikaen og hormonene man får i seg ved å spise kjøtt / meieriprodukter. Her er ikke norge noe annerledes.
- Å spise kjøtt er like kreftfremkallende som å røyke.
- Kun 6 – 10% av alle krefttyper kan overføres genetisk. Om så, så er det ikke en eller to kreftceller som er ille, men maten du spiser kan, og VIL booste disse cellene om du spiser feil (det vil si, for mye kjøtt). Som de fleste av oss gjør..
- Egg er såpass lite godt for deg at det nå i amerika ikke engang er lovlig å si i reklamer at det er sunt eller bra for deg med hele setninger. Så de må heller si "a good egg", det er ikke lovlig å si "good for YOU".
- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør. Dette fører til at man vil ha mer og mer, uansett om man går opp i vekt og huden tar skade.
- Det verste er at alt handler om kun en ting – penger. Apotekene skal tjene penger, regjeringen skal tjene penger, og kjøttindustrien skal. Disse tre jobber ofte sammen. Altså, bare SE den dokumentaren, den er så bra!

Det er trist at alt handler om penger. Både våre egne hoder, men samfunnet vi lever i."



Hentet 03.09.2020 fra innlegget «Åpne øynene», Sophie Elises blogg, 10. september 2017 ([https://sophieelise.blogg.no/1505064777\\_pne\\_ynene.html](https://sophieelise.blogg.no/1505064777_pne_ynene.html))

1.1 a) Bør man stole på informasjonen i kilde 1? \*

- Ja
- Nei

1.1 b) Hvilket av alternativene om kilde 1 er du mest enig i? \*

Ordforklaring: Troverdig = Noe man kan stole på.

- Den er troverdig, fordi forfatteren skriver på en måte jeg forstår.
- Den er troverdig, fordi forfatteren har undersøkt temaet nøye, og har gode begrunnelser.

- Den er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser
- Den er ikke troverdig, fordi alt bloggere skriver er skrevet kun for å selge produkter.

### Informasjon til de neste oppgavene

Et argument består av en påstand og en begrunnelse. I eksemplene på argumenter under er påstandene markert i **gult** og begrunnelsene i **grønt**.

«**Jeg mener at det er galt å fiske med krok**, fordi **vi vet at mange dyr kan føle smerte**, og fisk er et dyr.»

«**Du burde ikke stupe fra fjellet og ut i havet**, for **det gjorde broren min, og han brakk nesa**.»

I en tekst er **påstander** budskap i teksten. **Begrunnelser** brukes for å bygge opp troverdigheten til forfatterens budskap.

1.1 c) Hvilken av påstandene under prøver kilde 1 å formidle? \*

Ordforklaring: *Formidle = Det man prøver å si*

- Å spise kjøtt er like kreftfremkallende som å røyke
- Å spise kjøtt er sunt, så lenge man begrenser mengden.
- Å røyke er mindre kreftfremkallende enn mange tror.
- Hun har undersøkt temaet nøye, og viser til forskning.

1.1 d) Hvordan begrunner forfatteren i kilde 1 at hun funnet informasjon om temaet? \*

- Ost er bevist avhengighetsskapende og trigger (i mye mindre skala) de samme områdene i hjernen vår som heroin gjør.
- Hun søkte blant annet på google.
- Hun har undersøkt temaet ved å ta kontakt med en fagperson.
- Hun har undersøkt temaet nøye, og kan vise til forskning.

1.1 e) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.1 f) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 1 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 1.



- Enig
- Uenig

1.1 g) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 1 kommer fra en fagperson.

- Enig
- Uenig

1.1 h) Er du enig i hele utsagnet under? \*

Kilde 1 er troverdig, fordi påstandene stemmer med det jeg har lært på skolen om kjøtt og røyking fra før.

- Enig
- Uenig

1.1 i) Er du enig i hele utsagnet under? \*

Kilde 1 er ikke troverdig, fordi forfatteren har undersøkt temaet for lite, og derfor har mangelfulle begrunnelser.

- Enig
- Uenig

1.1 j) Er utsagnet under riktig hvis det slaktes 250 kuer i timen på et slakteri? \*

Da tar det 15 sekunder for en slakter å slakte en ku, fordi  $250\text{kuer} / 60\text{minutter} = 4\text{ kuer i minuttet}$ .

- Det er sant.
- Det er ikke nødvendigvis sant.



Sideskift

Side 5

## 1.2 Kilde 2

Teksten under er hentet fra en internettside.

Les og undersøk teksten, før du svarer på oppgavene under.

### «– Ikke like kreftfremkallende

*Faktisk.no tar kontakt videre med Kreftregisteret, som samler inn data og utarbeider statistikk om kreftforekomsten i Norge og driver med forskning. Direktør Giske Ursin bekrefter at det ikke er like kreftfremkallende å spise kjøtt som å røyke. Ursin tror hun vet hvor misforståelsen om at det er like kreftfremkallende å spise kjøtt som å røyke kommer fra:*



I 2015 klassifiserte Verdens helseorganisasjon (WHO) prosessert kjøtt (som for eksempel pølser) som sikkert kreftfremkallende, mens rødt kjøtt er klassifisert som sannsynlig kreftfremkallende. Dermed kommer prosessert kjøtt nå i samme klasse som røyk. At både prosessert kjøtt og røyk er klassifisert som sikkert kreftfremkallende, er ikke det samme som at risikoen er like stor. Graden av kreftfremkallenhet er ekstremt forskjellig.

Rødt, saltet og bearbeidet kjøtt kan øke risikoen for tarmkreft. Ifølge Kreftregisteret er det imidlertid bare en moderat andel av forekomsten av tarmkreft som kan tilskrives kosthold og livsstil. Ursin sier følgende: - For lungekreft kan derimot over 80 prosent tilskrives røyking. Direktøren understreker at Kreftregisteret støtter kreftfrådene fra WHO som anbefaler å begrense inntaket av rødt og bearbeidet kjøtt. Dette fordi noe av kreftforekomsten kan forebygges med godt kosthold og en aktiv livsstil.»

Hentet 03.09.2020 fra «Nei, å spise kjøtt er ikke like kreftfremkallende som å røyke», Eva Akerbæk, Mina Liavik Karlsen og Silje S. Skiphamn, 15.09.2017 (<https://www.faktisk.no/faktasjekker/BwD/nei-a-spise-kjott-er-ikke-like-kreftfremkallende-som-a-royke>)

1.2 a) Var informasjonen i kilde 2 troverdig? \*

- Ja
- Nei

1.2 b) Hva fortalte kilde 2 om kreftfaren med å spise rødt kjøtt? \*

- Det er like kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt som å røyke.
- Det er ikke kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det kan være kreftfremkallende å spise rødt, saltet og bearbeidet kjøtt.
- Det er kun bearbeidet kjøtt som er like kreftfremkallende som å røyke

1.2 c) Hvem har skrevet kilde 2? \*

- Giske Ursin
- En i kreftregisteret har skrevet den for faktisk.no.
- Det er to journalister som jobber i faktisk.no
- Det er tre journalister som jobber i faktisk.no
- Kreftregisteret
- To ansatte i kreftregisteret
- Tre ansatte i kreftregisteret
- Sophie Elise

1.2 d) Hva slags begrunnelse bruker kilde 2 for hvordan de kom frem til påstandene i teksten? \*

- Det er kun bearbejdet kjøtt som er like kreftfremkallende som å røyke
- Det begrunnes ved at faktisk.no har lest forskning på temaet.
- Faktisk.no tok kontakt med kreftregisteret, som forsker på temaet
- Det begrunnes ved at det er like kreftfremkallende å spise rødt, saltet og bearbejdet kjøtt som å røyke.

1.2 e) Er du enig i hele utsagnet under? \*

Kilde 2 er ikke troverdig, fordi forfatteren forsøker å selge et produkt, og har derfor en interessekonflikt.

- Enig
- Uenig

1.2 f) Er du enig i hele utsagnet under? \*

Forfatteren i kilde 2 begrunner påstandene sine med troverdige kilder og forskning. Dette øker troverdigheten til kilde 2.

- Enig
- Uenig

1.2 g) Er du enig i hele utsagnet under? \*

Informasjonen i kilde 2 kommer fra en fagperson, og dette gjør kilden troverdig.

- Enig
- Uenig

1.2 h) Er du enig i hele utsagnet under? \*

Påstandene i kilde 2 er mindre ekstreme enn i kilde 1, og dette øker troverdigheten til kilde 2

- Enig
- Uenig



### 1.3. Sammenligning av kilde 1 og 2


**Tips:**

- Husk at du kan bla tilbake i testen, men unngå å endre på oppgaver du allerede har besvart.

1.3 a) Hvilken kilde synes du var mest troverdig \*


- Kilde 1 (Sophie Elise «Åpne øynene»)
- Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»)

### 1.3 a.a Hva var de to viktigste faktorene når du synes kilde 1 var mest troverdig?

 Dette elementet vises kun dersom alternativet «Kilde 1 (Sophie Elise «Åpne øynene»))» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

### 1.3 a.b Hva var de to viktigste faktorene når du synes kilde 2 var mest troverdig?

 Dette elementet vises kun dersom alternativet «Kilde 2 (Faktisk.no «Ikke like kreftfremkallende»))» er valgt i spørsmålet «1.3 a) Hvilken kilde synes du var mest troverdig»

- Teksten var godt skrevet, og lett å forstå.
- Informasjonen i teksten kom fra en fagperson.
- Teksten hadde gode begrunnelser for påstandene sine.
- Påstandene i teksten stemte overens med det jeg visste fra før.
- Påstandene i teksten var basert på forskning.
- Jeg synes det var interessant å lese den.

### 1.3 b) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det er like kreftfremkallende å spise kjøtt som å røyke»? \*

- Ja
- Nei

### 1.3 c) Nå har du lest to kilder om kjøtt. Vil du si deg enig i påstanden «det kan være kreftfremkallende å spise kjøtt»? \*

- Ja
- Nei

### 1.3 d) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

*Å spise kjøtt er ikke bra for klimaet, og derfor bør man unngå det.*

- Argumentet er relevant
- Argumentet er ikke relevant

1.3 e) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

*Det er mange næringstoffer i kjøtt som mennesker trenger, og derfor er det sunt.*

- Argumentet er relevant
- Argumentet er ikke relevant

1.3 f) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

*Professor Birger Svihus ved NMBU mener at kjøtt inneholder viktige vitaminer og proteiner.*

- Argumentet er relevant
- Argumentet er ikke relevant

1.3 g) Vurder om argumentet under er relevant for kreftfaren ved å spise kjøtt: \*

*Ikke alle kjøttprodukter er bevist kreftfremkallende, som for eksempel kyllingfilet.*

- Argumentet er relevant
- Argumentet er ikke relevant

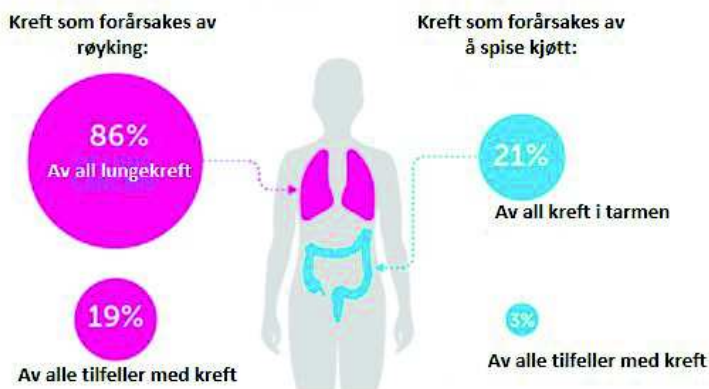


Sideskift

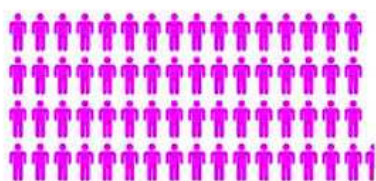
**Undersøk illustrasjonen under, og svar på oppgavene**

## Røyking sammenlignet med å spise kjøtt:

- Hvor stor er risikoen for å utvikle kreft?



Hvor mange får kreft av å røyke sammenlignet med å spise kjøtt:



Hvis ingen røyket ville det vært 64 500 færre tilfeller med kreft

 = 1000 tilfeller med kreft



Hvis ingen spiste kjøtt ville det vært 8 800 færre tilfeller med kreft

Tallene og illustrasjonen er hentet fra Cancer Research UK

1.4 a) Hvor mange mennesker ville unngått å få kreft hvis ingen røykte, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500
- 21%

1.4 b) Hvor mange mennesker ville unngått å få kreft hvis ingen spiste kjøtt, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500
- 21%

1.4 c) Hvor stor andel av all lungekreft kommer av røyking, ifølge illustrasjonen? \*

- 86%
- 8 800
- 64 500

21%

3%

1.4 d) Hvor stor andel av alle krefttilfeller skyldes av at folk spiser kjøtt, ifølge illustrasjonen? \*

21%

19 %

64 500

86 %

3 %

1.4 e) Hva betyr hver rosa mann i illustrasjonen? \*

Hver rosa mann viser ett tilfelle av kreft som skyldes at man spiser kjøtt.

Hver rosa mann viser ett tilfelle av kreft som skyldes røyking.

Hver rosa mann viser tusen tilfeller av kreft som skyldes at de spiser kjøtt.

Hver rosa mann viser tusen tilfeller av kreft som skyldes røyking.

Hver rosa mann viser ett liv som ville blitt spart hvis ingen spiste kjøtt.

Hver rosa mann viser ett liv som ville blitt spart hvis ingen røykte.



Sideskift

Side 8

## Del 2 av testen: Å kunne trekke konklusjoner



I denne serien av oppgaver vil du stegvis bli presentert for ny informasjon om hvorfor **tang og tare forsvinner fra havbunnen** utenfor kysten ved Alaska.

Målet er at du ut fra denne informasjonen skal finne **den mest sannsynlige forklaringen** på **hvorfor** det skjer.

### 2.1 Bruk informasjonen, og vurder forklaringen:

- Tang og tare er sjøplanter som lever på bunnen av havet.
- På havbunnen utenfor Alaska har det gjennom flere år blitt observert at det blir mindre tang og tare. Forskerne er bekymret for at denne nedgangen kan få store konsekvenser.
- Havtemperaturen har sakte, men sikkert, økt grunnet global oppvarming.

2.1 a) Er du enig i at konklusjonen under kan stemme? \*

**Konklusjon:** Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

- Ja
- Nei

2.1 b) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

2.1 c) Hvilken annen forklaring kan du se for deg for hvorfor tang og tare forsvinner? \*



### 2.2 Bruk informasjonen du har fått til å finne forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen.

- En forklaring er mulig, hvis den ikke utelukkes fra informasjonen du får.
- Husk at du kan bla tilbake for å se tidligere informasjon. Det er viktig at du ikke endrer på svarene du allerede har gitt på tidligere oppgaver.

#### Ny informasjon:

- Antallet kråkeboller har gått opp i samme tidsrom som det har blitt mindre tang og tare.
- Sjøotere er et rovdyr, og spiser kun byttedyr.
- Kråkebollen er et byttedyr.
- Tidligere var det et problem mennesker jaktet på sjøotere for å bruke pelsen deres til å lage eksklusive pelsprodukter som pelskåper. Dette førte til at sjøoterne nesten ble utryddet rundt Alaska på starten av 1900-tallet.
- Et byttedyr er maten til rovdyr. Hvis byttedyret forsvinner vil rovdyret enten dø av sult eller måtte spise en annen art.
- Spekkhoggere er en hval, og et rovdyr. Den spiser for det meste fisk. Den har ingen naturlige fiender.

2.2 a) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*



- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

2.2 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

2.2 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

2.2 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

2.2 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

2.2 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt igjen de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

Dette er en mulig forklaring.

Dette er ikke en mulig forklaring

### 2.3 Du skal vurdere de samme forklaringene som på del 3.2, men du får mer informasjon som du kan bruke.

Husk at du kan bla tilbake

for å se på tidligere informasjon. Det er viktig at du ikke endrer svarene du allerede har gitt på tidligere oppgaver.

#### Ny informasjon:

- Kråkeboller er planteetere.
- Et økosystem består av mange arter (altså ulike typer dyr, planter og bakterier), som påvirker hverandre.
- Det ble ulovlig å jakte på sjøotere i 1911, fordi sjøoterne var på vei til å bli utryddet.
- Mange arter er sårbare for endringer i økosystemet, som forandringer i temperatur eller hvis det kommer nye ukjente arter inn i økosystemet.
- Mennesker påvirker økosystemer blant annet gjennom global oppvarming, forsøpling av havene og ved å overfiske.
- Hvis mennesker fortsetter å kaste søppel i havene vil det i 2050 være mer plastikk i havet enn fisk.
- Tang og tare er tilhører planteriket, og er ikke et dyr.
- Sportsfiskere klager på at det har blitt vanskeligere å få fisk når de fisker.
- I 1970 var det like mange sjøotere i Alaska som det var før menneskene begynte å jakte på den. Å forby jakt på sjøotere er en suksess.

2.3 a) Har du nok informasjon til å være sikker på hva som er forklaringen på hvorfor det har blitt mindre tang og tare på havbunnen? \*

- Ja
- Nei

2.3 b) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 c) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen. \*

- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 d) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 e) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 f) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring

2.3 g) Vurder om forklaringen under kan forklare hvorfor det har blitt mindre tang og tare på havbunnen \*

- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.
- Dette er en mulig forklaring.
- Dette er ikke en mulig forklaring



Sideskift

Side 11

**2.4 Velg den mest sannsynlige forklaring for hvorfor det har blitt mindre tang og tare på havbunnen utenfor Alaska.**

- Husk at du kan bla tilbake for å se på tidligere informasjon.

**Ny informasjon:**

- En spekkhogger kan spise 1825 sjøotere i året, hvis den bare spiser otere.
- Man kan finne samme type tang som finnes på havbunnen utenfor Alaska, i områder lenger syd, hvor det er et varmere klima.
- Det har blitt 25% færre sjøotere i året på havbunnen utenfor Alaska siden 1990-tallet.
- Forskerne er forundret over at de ikke har funnet flere døde sjøotere på strender. Dette gjør at de utelukker sykdom som hovedårsaken til hvorfor det har blitt færre sjøotere, fordi syke sjøotere ville blitt skylt opp på strendene når de døde.

2.4 a) Har du nok informasjon til å utelukke de fleste forklaringene om hvorfor det har blitt mindre tang og tare? \*

- Ja
- Nei

2.4 b) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

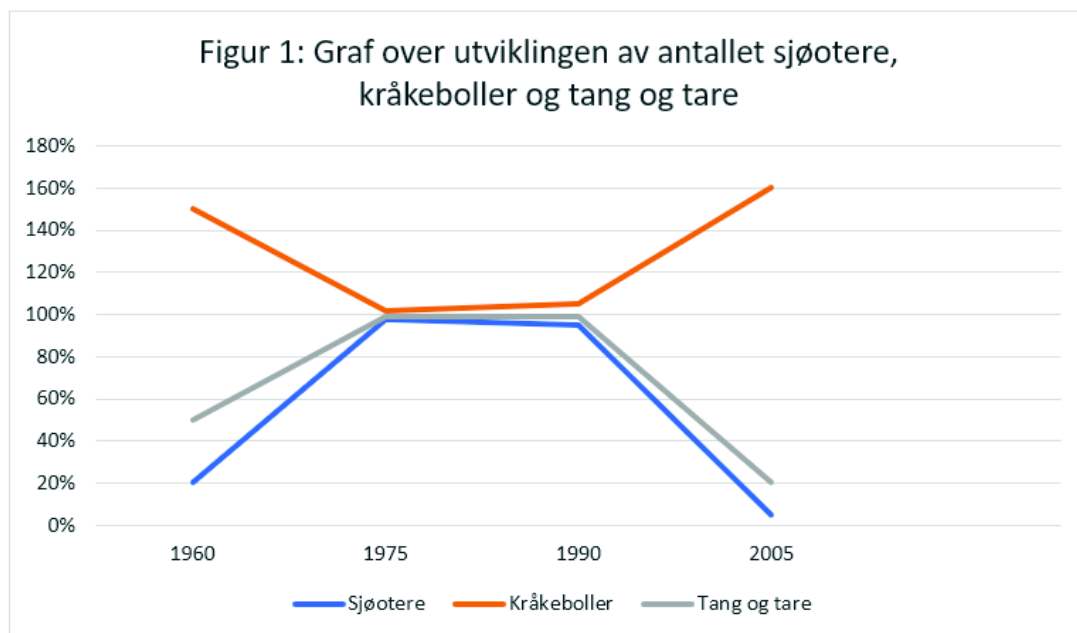
- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.

2.4 c) Hvordan tenkte du for å komme frem til den mest sannsynlige forklaringen? \*

 Sideskift

Side 12

2.5 Undersøk grafen under, og svar på spørsmålene



Figur 1: grafen gir en oversikt over andelen av arten i prosent, i forhold til hvordan det var før mennesker påvirket økosystemet. Verdier over eller under 100% viser at det har blitt flere eller færre av arten.

2.5 a) Hva er viser figur 1? \*

- Figur 1 viser en oversikt over temperaturer i havområdet utenfor Alaska fra 1960 til 2005.
- Figur 1 viser en oversikt over temperaturer i havområdet utenfor Alaska fra 1960 til i dag.
- Figur 1 viser en graf over utviklingen av antallet sjøotere, kråkeboller og tang og tare utenfor Alaska, i tidsperioden 1960 til i dag.
- Figur 1 viser en graf over utviklingen av antallet sjøotere, kråkeboller og tang og tare utenfor Alaska, i tidsperioden 1960 til 2005.

### 2.5 b) Hva kan tolkes ut fra grafen i figur 1? \*

- Tabellen viser at når det blir flere kråkeboller, blir det flere sjøotere og mindre tang og tare.
- Tabellen viser at nedgangen i antallet sjøotere, skjer samtidig som det blir flere kråkeboller og mindre tang og tare.
- Tabellen viser at når det blir mindre tang og tare, fører dette til at det blir det flere kråkeboller.
- Tabellen viser at når det blir færre kråkeboller, fører dette til at det blir flere sjøotere.

### 2.5 c) Hvilken av forklaringene nedenfor er den mest sannsynlige forklaringen for hvorfor det har blitt mindre tang og tare på havbunnen? \*

Bruk informasjonen fra grafen til å svare på spørsmålet.

Hvis du var sikker på forklaringen din på oppgave 3.4 b), kan du svare det samme som du gjorde der.

- Tang og tare tåler ikke at havet blir varmere, fordi den er sensitiv for endringer av temperatur.
- Sjøotere spiser tang og tare, og dette forklarer nedgangen, fordi det har blitt flere sjøotere etter at det ble ulovlig å jakte på dem.
- Plastikk fører til at det har blitt færre fisk. Dette har ført til at spekkhoggeren har begynt å spise tang og tare.
- Det blir for mange kråkeboller som spiser tang og tare, fordi spekkhoggerne har endret matvaner fra fisk til sjøotere.
- Det kan ha blitt færre sjøotere grunnet sykdom eller pest, som har ført til det blir mange kråkeboller, som spiser tang og tare.
- Forbruket av oterpels har økt de siste årene, og ført til jakt på sjøotere. Dette har ført til at det har blitt flere kråkeboller som spiser tang og tare.



Sideskift

Hvor enig er du i påstandene?

Spørsmålene har ikke et fasitsvar. Svar ut fra hvor mye du kjenner deg igjen i påstandene.

	Helt enig	Litt enig	Verken enig eller uenig	Litt uenig	Helt uenig
Å lære om nye temaer er spennende. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alle har rett på sin mening. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er flink til å finne kreative løsninger på oppgaver. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det viktigste i en diskusjon er å vinne. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg samler mye informasjon før jeg tar en avgjørelse *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er nysgjerrig på det vi lærer i naturfag. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Å lære å tenke kritisk er det noe av det viktigste man lærer på skolen. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ofta legger jeg lite innsats i å lære noe nytt på skolen. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg viser respekt for andres meninger. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg får en følelse av panikk når oppgaver er vanskelige *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Uansett hva som skjer bør man holde på verdiene sine, og ikke revurdere dem. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I stedet for å jobbe grundig med å løse en oppgave, spør jeg en klassekamerat om svaret. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg synes det man lærer i naturfag er viktig å kunne når man skal delta i demokratiske valg. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg har lært mye hvordan man skal tenke kritisk på skolen. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg bruker mye tid på å lese om ting jeg er nysgjerrig på. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg er åpen for ny informasjon, selv om jeg er ganske sikker i saken. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Jeg gir meg når oppgavene er vanskelige *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Det er viktig for meg at den kunnskapen jeg har er så sikker som mulig. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg vurderer konsekvensene før jeg tar en avgjørelse. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
På fritiden tenker jeg over ting jeg lærer i naturfag. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
På skolen har jeg lært hvordan man kan være kildekritisk. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg liker å lese om temaer jeg vet lite om. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg liker å lese på internett om meninger jeg er uenig med, for å få innsikt i hvordan andre tenker. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg reflekterer godt. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg velger ut informasjon som passer med det jeg mener. *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jeg bruker flere kilder når jeg undersøker et tema *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Naturfagskunnskapene jeg har hjelper meg med å være kildekritisk *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

#### Vedlegg 4.4: Observasjonsskjema til faglærerne i studie 4

Antall elever i totalt gruppen _____	Antall elever deltar _____	Frafall (de som velger å ikke delta) _____	Vanskelige ord det ble spurt om:
<b>Del 1</b> <b>Kildekritikk:</b>	Antall spørsmål om del <b>1.0 og 1.1</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>1.2</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>1.3</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>1.4</b>	Oppgaver/instrukser det ble spurt om:	
<b>Del 2</b> <b>Konklusjoner</b>	Antall spørsmål om del <b>2.1</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>2.2</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>2.3</b>	Oppgaver/instrukser det ble spurt om:	
	Antall spørsmål om del <b>2.4</b>	Oppgaver/instrukser det ble spurt om:	
<b>Andre observasjoner / faktorer som er verdt å merke seg:</b>			



## Vedlegg 4.5: Resultater reliabilitetsanalyse på holdningskategoriene i studie 4

Holdningskategori	Cronbach alfa	Variabelnavn	Gj. Snitt	Cronbach alfa Hvis slettet	Forslag til endring	Revidert i studie 3
<b>Intellektuell nysgjerrighet</b> (Inquisitiveness)	0,52	I-1	4,38	0,39	Å lære om nye temaer er spennende.	X
		I-2	4,40	<b>0,68</b>	Ofte legger jeg lite innsats i å lære noe nytt på skolen. (negativ)	X
		I-3	3,69	0,36	Jeg bruker mye tid på å lese om ting jeg er nysgjerrig på.	
		I-4	3,36	0,33	Jeg liker å lese om temaer jeg vet lite om.	X
<b>Åpenhet for ny informasjon</b> (Open-Mindedness)	0,66	A-1	4,04	0,64	Alle har rett på sin mening.	X
		A-2	3,31	0,56	Jeg viser respekt for andres meninger.	X
		A-3	4,39	0,48	Jeg er åpen for ny informasjon, selv om jeg er ganske sikker i saken.	
		A-4	2,56	<b>0,70</b>	Jeg liker å lese på internett om meninger jeg er uenig med, for å få innsikt i hvordan andre tenker. ***	
<b>Selvtillit</b> (Critical Thinking Self-Confidence og Analyticity)	0,38	Se-1	2,57	0,33	Jeg er flink til å finne kreative løsninger på oppgaver.	X
		Se-2	3,53	0,25	Jeg får en følelse av panikk når oppgaver er vanskelige (negativ)	X
		Se-3	3,64	0,19	Jeg gir meg når oppgavene er vanskelige (negativ)	X
		Se-4	4,19	0,42	Jeg reflekterer godt.	X
<b>Sannhetssøkende</b> (Truth-Seeking)	0,3	Sa-1	3,36	-0,06	Det viktigste i en diskusjon er å vinne. (negativ) **	
		Sa-2	4,17	0,04	Uansett hva som skjer bør man holde på verdiene sine, og ikke revurdere dem. (negativ)	

		Sa-3	4,03	<b>0,60</b>	Det er viktig for meg at den kunnskapen jeg har er så sikker som mulig.	
		Sa-4	3,57	0,08	Jeg velger ut informasjon som passer med det jeg mener. (negativ) ** (***)	
<b>Betenksomhet</b> ovenfor avgjørelser  (Systematicity og Maturity)	0,38	B-1	3,88	<b>0,43</b>	Jeg samler mye informasjon før jeg tar en avgjørelse	X
		B-2	3,88	0,31	I stedet for å jobbe grundig med å løse en oppgave, spør jeg en klassekamerat om svaret. (negativ) **	
		B-3	2,36	0,30	Jeg vurderer konsekvensene før jeg tar en avgjørelse.	X
		B-4	4,25	0,19	Jeg bruker flere kilder når jeg undersøker et tema	X
<b>Holdninger til naturfag</b>	0,70	N-1	4,03	0,59	Jeg er nysgjerrig på det vi lærer i naturfag.	
		N-2	3,64	0,68	Jeg synes det man lærer i naturfag er viktig å kunne når man skal delta i demokratiske valg.	
		N-3	3,10	0,65	På fritiden tenker jeg over ting jeg lærer i naturfag.	
		N-4	4,03	0,63	Naturfagskunnskapene jeg har hjelper meg med å være kildekritisk. ***	
<b>Holdninger til viktigheten av kritisk tenkning i skolen</b>	0,68	H-1	4,06	0,65	Å lære å tenke kritisk er det noe av det viktigste man lærer på skolen.	
		H-2	3,72	0,62	Jeg har lært mye hvordan man skal tenke kritisk på skolen.	
		H-3	3,94	0,48	På skolen har jeg lært hvordan man kan være kildekritisk. ***	