



Norges miljø- og  
biovitenskapelige  
universitet

**Bacheloroppgave 2024 15 stp**

NMBU Veterinærhøgskolen  
Karianne Muri

## **En studie av inter- og intra- observatør-reliabilitet ved kvalitativ adferdsvurdering av katt**

A study of inter- and intra-observer-reliability of  
qualitative behaviour assessment of cats

**Syver Vik Enger, Tobias Humlekjær Palmgren og  
Sarah Wilberg**

Bachelor Dyrepleie  
Institutt for Produksjonsdyrmedisin

## Innholdsfortegnelse

1. Forord .....	4
2. Sammendrag.....	4
3. Forkortelser og definisjoner .....	4
4. Innledning.....	5
4.1 Dyrevelferd .....	5
4.1.1 Definisjoner av dyrevelferd .....	5
4.1.2 Dyrevelferd, livskvalitet, velvære og adferd .....	7
4.1.3 Målemetoder for dyrevelferd .....	7
4.2 Qualitative Behaviour Assessment – QBA .....	8
4.2.1 Bakgrunn og historie.....	8
4.2.2 Beskrivelse av metoden .....	9
4.2.3 Reliabilitet .....	10
4.2.4 QBA i forskning .....	10
4.2.5 QBA på katt .....	11
4.3 Bakgrunn for kattens adferd.....	11
4.3.1 Domestisering.....	11
4.3.2 Avl .....	12
4.3.3 Katteadferd i faglitteraturen .....	13
4.3.4 Shelterkatter .....	13
4.4 Formål.....	13
5. Materiale og Metoder .....	14
5.1 Gruppeinndeling.....	14
5.2 Kvalitative adferdsuttrykk.....	15
5.3 Visuell Analog Skala (VAS) .....	15
5.4 Gjennomføring av studien.....	15
5.4.1 Videoscoring.....	16
5.4.2 Kalibrering av observatørene .....	16
5.4.3 Feltscoring .....	17
5.5 Databehandling og statistikk .....	17
5.5.1 Gjennomsnitt og variasjonsbredde .....	17
5.5.2 Kendall's <i>W</i> .....	17
5.6 Litteratursøk .....	18
6. Resultater .....	19
6.1 Gjennomsnitt og variasjonsbredde .....	19
6.2 Delmål 1: Inter-observatør-reliabilitet.....	20
6.1.1 Videoscoring .....	20
6.1.2 Feltscoring .....	21
6.1.3 Oppsummering.....	21
6.3 Delmål 2: Endring i samsvar over tid.....	22
6.4 Delmål 3: Intra-observatør-reliabilitet.....	22

7. Drøfting.....	24
7.1 Drøfting av funn.....	24
7.1.1 Gjennomsnitt og variasjonsbredde .....	24
7.1.2 Delmål 1: Inter-observatør-reliabilitet.....	24
7.1.3 Delmål 2: Observer-drift og effekt av feltkalibrering.....	25
7.1.4 Delmål 3: Intra-observatør-reliabilitet.....	25
7.2 Sammenligning med tidligere studier .....	26
7.3 Feilkilder .....	27
7.3.1 Målinger utført for hånd.....	27
7.3.2 Kalibreringsprosessen .....	27
7.3.3 Mangel på variert adferd .....	27
7.3.4 Effekt av kattens miljø på observatøren .....	28
7.3.5 Tidsbegrensning: 2 minutter per katt .....	28
7.3.6 Eksponering for stimuli versus ren observasjon.....	29
7.3.7 Minimum og maksimum.....	30
7.3.8 Tap av fokus under scoring .....	30
7.4 Diskusjon av uttrykk .....	31
7.4.1 Motpoler.....	31
7.4.2 Sosiabilitet .....	31
7.4.3 utfordringer knyttet til "Fixed List" .....	31
7.5 Fordeler og ulemper med metoden .....	32
7.5.1 Fordeler ved bruk av QBA.....	32
7.5.2 Ulemper ved bruk av QBA .....	33
7.6 Bruksområder for metoden.....	34
7.6.1 Shelterprotokoll for katt .....	34
7.6.2 Identifisering av negativ velferd og terapibehov .....	34
7.6.3 Tilsyn med dyrevelferd .....	34
7.6.4 Adopterbarhet .....	35
7.6.5 QBA-app: for en mer tilgjengelig metode .....	35
8. Konklusjon .....	36
9. Veien videre.....	37
10. Takk til bidragsytere.....	38
11. Summary .....	38
12. Referanser.....	39
13. Vedlegg .....	42

# 1. Forord

Denne oppgaven er Bachelorgradsarbeidet til 3 dyrepleierstudenter ved Norges miljø- og biovitenskapelige universitet (NMBU) våren 2024. Oppgaven presenterer en studie i inter- og intrareliabilitet mellom observatører i bruk av verktøyet Qualitative Behaviour Assessment (QBA) på katter i sheltere hos 2 dyrevernsorganisasjoner på Østlandet. Verktøyet er blitt brukt både under indirekte observasjon av katter på video og under direkte observasjon av katter i shelterne.

# 2. Sammendrag

**Tittel:** “En studie av inter- og intra-observatør-reliabilitet ved kvalitativ adferdsvurdering av katt”

**Forfattere:** Syver Vik Enger, Tobias Humlekjær Palmgren & Sarah Wilberg

**Veileder:** Karianne Muri, førsteamanuensis II, Institutt for produksjonsdyrmedisin, NMBU veterinærhøgskolen

**Medveiledere:** Randi Oppermann Moe, professor, Institutt for produksjonsdyrmedisin, NMBU veterinærhøgskolen & Solveig Marie Stubsjøen, seniorforsker, Seksjon husdyr, vilt og velferd, Veterinærinstituttet

I denne studien undersøkte tre dyrepleierstudenter inter- og intra-observatør-reliabilitet ved bruk av metoden Qualitative Behaviour Assessment (QBA) på katt. Studien brukte en forhåndsdefinert liste med 20 adferdsuttrykk utarbeidet av veilederne, og kattene ble scoret på en visuell analog skala (VAS) på 125mm. Kattene som ble observert var omplasseringskatter oppstallet hos dyrevernsorganisasjoner. Videoklipp av kattene ble scoret to ganger med 4 måneders mellomrom. En kontrollgruppe på 3 medstudenter scoret også videoene. Inter- og intra-observatør-reliabilitet ble undersøkt ved bruk av Kendall's *W*, som gir et mål på samsvar.

Generelt var inter-observatør-reliabilitet moderat-til-høyt for de aller fleste adferdsuttrykkene, uavhengig av gruppeinndeling. De negativt ladede uttrykkene hadde generelt noe lavere samsvar enn de positivt ladede. Intra-observatør-reliabilitet for hver observatør var generelt høyt for alle uttrykkene. Det ble også gjennomført scoring av 20 katter i felt hos organisasjonene. Feltscoreningene viste noe lavere inter-observatør-reliabilitet i felt enn på video, men også her var det moderat-til-høyt for de fleste uttrykk.

# 3. Forkortelser og Definisjoner

QBA	Qualitative Behaviour Assessment. Kvalitativ adferdsvurdering
VAS	Visuell Analog Skala
FOD	Foreningen for Omplassering av Dyr
DOOA	Dyrebeskyttelsen Norge Oslo og Akershus
SRUC	Scotland's Rural College
FCP	Free Choice Profiling. En metode der observatører beskriver det de observerer med selvvalgte adjektiver
FL	Fixed List. En liste med forhåndsbestemte adferdsuttrykk

Inter-observatør-reliabilitet	Påliteligheten til en metode innad i en bestemt gruppe observatører
Intra-observatør-reliabilitet	Påliteligheten til en metode når brukt av samme observatør flere ganger over tid
Observer-drift	En observatørs gradvise endring i vurdering og scoring over tid
Kendall's <i>W</i>	Kendall's Coefficient of Concordance. Et statistisk mål på samsvar mellom observasjoner og/eller observatører
Shelter	Et lokale for midlertidig oppstalling av hjemløse dyr
Feltobservasjon / i felt	Direkte observasjon av enkeltdyr

## 4. Innledning

Denne studien undersøker reliabilitet av kvalitativ adferdsvurdering på katt. Studentene som deltok benyttet metoden ved observasjon av shelterkatter på video og i felt, og resultatene av disse observasjonene ble samlet og analysert. Målet med slike reliabilitetsstudier er å bidra til å undersøke om metoden er pålitelig til å vurdere dyrevelferd.

### 4.1 Dyrevelferd

#### 4.1.1 Definisjoner av dyrevelferd

Begrepet "dyrevelferd" har blitt definert ulikt gjennom tidene og i ulike fagmiljø (Deane & Valentine, 2022; Duncan, 2020; Foreningen Norske Etologer, u.å.; Wemelsfelder, 1997; Yeates, 2022). Slik sett fremstår det som et dynamisk begrep med en pågående diskurs.

Dyrevelferd dreier seg om hvorvidt dyrene har det bra. Diskursen dreier seg om hva dyrene trenger for virkelig å ha det bra, og om hvordan vi kan vurdere om det enkelte dyret selv opplever å ha det bra og å ha et godt liv.

Vann, mat og ly – det forsto vi mennesker tidlig at dyrene våre hadde behov for. Likeledes vern mot rovdyr, skade eller sykdom. Uten disse fasiliteter ville vi selv heller ikke ha nytte av dyreholdet.

#### De 5 friheter og de 5 domener

På 1960-tallet fikk Roger Brambell i oppdrag å organisere og lede en omfattende utredning av levekår hos produksjonsdyr i Storbritannia (Duncan, 2020; Fraser, 2008). Utredningen påviste store problemer i dyrenes levekår samt mange konkrete endringsbehov. I sin rapport konkretiserte Brambellkommisjonen blant annet at alle dyr skal ha frihet til å reise seg, frihet til å legge seg, frihet til å snu seg, frihet til å strekke bena og frihet til å stille seg. Denne listen fikk tilnavnet "Brambells 5 friheter" (Keeling et al., 2018; Michigan State University, u.å.).

I etterkant av Brambellkommisjonens rapport ble det nedsatt en nasjonal komité for videre oppfølging av dyrevelferd hos britiske produksjonsdyr, og i 1979 ble den nye listen over 5 friheter presentert som kriterium for god dyrevelferd:

- Frihet fra sult og tørste
- Frihet fra ubehag
- Frihet fra smerte, skade og sykdom
- Frihet til å utøve normal adferd
- Frihet fra frykt og stress

Etter hvert har disse 5 friheter blitt videreutviklet til konseptet om "De 5 domener" (Deane & Valentine, 2022):

- Ernæring
- Omgivelser
- Helse
- Måltrettet adferd (Behavioral interactions)
- Mental tilstand / opplevelser

I faglitteraturen og forskning er de ulike aspektene av de 5 friheter og de 5 domener stadig under drøfting og utvikling (Mellor & Beausoleil, 2015; Mellor, 2017; Mellor et al., 2020). Nye spesifiseringer og forståelsesmodeller kommer til, og likeledes nye verktøy.

Dyrevelferd er et fokus og interessefelt for mange ulike fagmiljø (veterinærmedisin, etologi, filosofi, etikk, biologi, samt landbruksnæring og tilsynsmyndigheter for dyrevelferd) som naturlig nok bringer inn ulike vinklinger (Fraser et al., 1997; Fraser, 1999; Fraser, 2008; Norges forskningsråd, 2005).

Foreningen Norske Etologer (FNE) skriver på sine nettsider (Foreningen Norske Etologer, u.å.) at «Dyrevelferd handler om hvordan dyret har det» og at den mest kjente definisjonen på dyrevelferd ble forfattet av professor Donald M. Broom i 1986: «Dyrevelferd er dyrets tilstand med hensyn på å mestre sitt miljø». FNE utdyper at «denne definisjonen av dyrevelferd handler om hvorvidt dyret har det bra, nøytralt eller dårlig, og at dyrets tilstand er et resultat av hvordan det takler miljøet det lever i» (Foreningen Norske Etologer, u.å.).

David Fraser, zoolog og professor i dyrevitenskap skriver at dyrevelferd fordrer at (minst) 3 etiske hensyn oppfylles:

- 1) å leve et naturlig liv gjennom utvikling og bruk av dyrenes naturlige evner
  - 2) at dyrene skal oppleve velvære ved å være fri for vedvarende eller sterk frykt, smerte eller andre negative tilstander, og skal erfare normale gleder
  - 3) tilfredstillende helse og vekst samt normal fysiologisk og adferdsmessig funksjon.
- (Fraser et al., 1997)

### 4.1.2 Dyrevelferd, livskvalitet, velvære og adferd

I dagligtale og media presenteres “dyrevelferd” gjerne som udelt positivt og at fraværet av, eller reduksjonen i, velferden er det negative. Norske fagmiljø innen veterinærmedisin og landbruk bruker gjerne begrepet “dyrevelferd” som en nøytral betegnelse, i at dyrevelferden kan være enten god eller dårlig (Moe, 2016; Norges forskningsråd, 2005).

Begrepet “well-being” (på norsk “velvære”) benyttes ofte i engelskspråklig litteratur om dyrevelferd (Appleby & Sandoe, 2002; Appleby et al., 2018; McMillan & Yeates, 2020). Dette anser vi å være en verdifull vinkling og presisering, da det retter fokus direkte på dyrenes opplevelser og emosjoner framfor menneskenes tilrettelegging av deres levekår. Er dyret redd? Kjeder det seg? Er dyret frustrert eller deprimeret? Har det smerter eller er det ukomfortabelt trass i menneskenes forsøk på å oppfylle alle friheter og domener?

Dyrenes opplevelse av velvære vil logisk sett være en viktig indikator på om dyrevelferden er god. Utfordringen ligger i å tolke dyrenes opplevelse av velvære eller mangel på velvære.

Rapporten “*Forskningsbehov innen Dyrevelferd i Norge*” fra Norges forskningsråd i 2005 definerer dyrevelferd som “individets subjektive opplevelse av sin mentale og fysiske tilstand som følge av dets forsøk på å mestre sitt miljø” (Norges forskningsråd, 2005)

### 4.1.3 Målemetoder for dyrevelferd

Ulike kartleggings- og evalueringsverktøy er blitt utviklet for bruk på ulike dyrearter og ulike typer dyrehold. For å registrere aspekter ved dyrs velferd brukes eksempelvis ofte et etogram: en liste over viktige adferder man er interessert i å registrere for den enkelte art, med tilhørende beskrivelse av kroppsspråk og fysiske aktiviteter. (Universities of Glasgow & Edinburgh Napier, 2015) Et etogram for katt kan beskrive frykt med at øynene er vidåpne, ørene flate, og at katten ikke vil spise eller drikke. Etogrammer inneholder gjerne illustrasjoner.

Mye av fokuset innenfor dyrevelferdsvurdering har ligget på å fange opp problemer eller udekkede behov knyttet til dyrenes fysiske helse, adferd og miljø, og da gjerne i forhold til definerte eller vedtatte retningslinjer og/eller forskrifter for den enkelte dyreart og dyreholdets driftsform (Butterworth et al., 2018). Slike vurderingsmetoder innebærer registrering av kvantitative data (for eksempel areal per dyr, temperatur, luftkvalitet), data om dyrenes fysiske omgivelser (for eksempel underlag, skjulesteder) og ressurstilgang (som mat og vann) (Norges forskningsråd, 2005). Alle disse er eksempler på ressursbaserte velferdsindikatorer. Disse er viktige og verdifulle for å sikre at viktige retningslinjer eller forskrifter for dyrehold følges.

Andre (ofte supplerende) metoder for velferdsvurdering ser på dyrebaserte velferdsindikatorer, ved å benytte seg av kvantitativ og/eller kvalitativ registrering av dyrenes adferd og kroppsspråk. Begrepet “kroppsspråk” betegner at observerte bevegelser eller kroppstillinger kan uttrykke og dermed gi informasjon om sinnstilstand, fokus, motivasjon og emosjoner (Braastad, 2019; Leyhausen, 1979).

Da ulike dyrearter har ulike egenskaper, behov, bevegelsesmønstre og derved kroppsspråk, er det nærliggende å anta at tolkning av kroppsspråket vil fordre både kunnskap om og erfaring med arten. Registrering av sykdom- og skadeinsidens er også vanlig. Kvantitative metoder for vurdering og registrering av dyrevelferd følger gjerne standardiserte protokoller (for eksempel etogrammer), utarbeidet for den enkelte dyreart.

Dyrebasert tilnærming er fokuset i EUs Welfare Quality prosjekt, hvis mål var å forbedre dyrevelferden hos produksjonsdyr i Europa. (PAW, u.å.) Prosjektet resulterte i utvikling av omfattende velferdsprotokoller for ulike arter, som tar i bruk en rekke metoder som samlet sett skal gi en helhetlig vurdering av dyrenes velferd. Vurderingsmetodene i disse protokollene baserer seg i størst mulig grad på dyrebaserte velferdsindikatorer. For eksempel ser man på om dyret har sår eller skader på kroppen, og om dyret utfører normal sosial eller undersøkende adferd. Enkelte indikatorer er likevel ressursbaserte der det ikke finnes et godt dyrebasert alternativ. (Welfare Quality Network, u.å.) Basert på Welfare Quality prosjektet har det også blitt laget en tilsvarende velferdsprotokoll for hunder på shelter. (Arena et al., 2014) I Welfare Quality og Shelter Quality protokollene deles velferdsvurderingen inn i 4 hovedkategorier (fôr, oppstalling, helse og adferd) som hver består av 3-4 kriterier, med diverse delkriterier som kan vurderes og registreres. Blant delkriteriene under adferd for hund finner man blant annet unormal adferd, respons på mennesker og emosjonell status. (Arena et al., 2014)

For å undersøke velferdsriteriet som handler om dyrs emosjonelle status, eller "positive emotional state", behøver man en metode som forsøker å avdekke dyrets opplevelse av sin egen situasjon. Her kommer metoden "Qualitative Behaviour Assessment" (QBA) inn i bildet.

## **4.2 Qualitative Behaviour Assessment (QBA)**

### **4.2.1 Bakgrunn og Historie**

Grunnlaget for metoden ble lagt av Francoise Wemelsfelder ved Scottish Agricultural College i 1997. I sin artikkel "*The scientific validity of subjective concepts in models of animal welfare*" (Wemelsfelder, 1997), beskriver hun behovet for en metode som vurderer velferd basert på dyrs subjektive opplevelse av sin egen situasjon.

Wemelsfelders motivasjon handlet i stor grad om hennes filosofiske standpunkt. Hun ønsket å utfordre ettervirkningene av de langstående prinsippene angående dyresyn som den franske filosofen René Descartes formulerte på 1600-tallet, innenfor vitenskapens rammer. Hun pekte på at Descartes så på dyr som objekter uten en sjel som kun fulgte mekanistiske prinsipper og som ikke hadde følelser, og mente på det tidspunktet (1997) at forskningen på dyrevelferd fremdeles hang igjen i et syn på dyrs subjektive opplevelse som noe som ikke kan måles. (Wemelsfelder, 1997) SNL skriver følgende om Descartes' lære:

"I mennesket er bevissthet eller sjel forent med et materielt og utstrakt legeme til ett vesen. (...) Dyrene har ingen tenkning og ingen sjel og er derfor bare legemer eller «maskiner»." (Tranøy & Ore, 2020)



Kartesiansk dualisme som filosofi innebærer altså et syn på (ikke-menneskelige) dyr som kompliserte mekanistiske objekter, hvis eneste funksjon er å respondere på stimuli fra omverdenen etter gitte mekanistiske lover. Subjektiv opplevelse, som her stammer fra en sjel som finnes utenfor det fysiske plan, oppfattes som noe unåelig for andre enn en selv, og noe som ikke tilskrives dyr. (Tranøy & Ore, 2020) Dette synet beskriver Wemelsfelder som lite treffende i dagens moderne samfunn. Tvert imot mener hun at vi mennesker, når vi interagerer direkte med dyr, oppfatter dem som subjekter med sitt eget individuelle perspektiv på virkeligheten rundt seg. Hun mener videre at mennesker har en evne til å analysere og beskrive dyrets adferd ut ifra en forestilling om "hvordan-det-er-å-være" dyret. Behovet for å utforske dette perspektivet vitenskapelig, og å forstå adferd ut ifra subjektets helhet, ikke som summen av kausale faktorer, ble grunnlaget for utforming av QBA-metoden. (Wemelsfelder, 1997) De neste 10 årene arbeidet Wemelsfelder og hennes kolleger med å utvikle en ny metode for kvalitativ velferdsvurdering av dyr. (Wemelsfelder, 2007).

#### **4.2.2 Beskrivelse av metoden**

Resultatet av Wemelsfelders arbeid ble en ny metode for vurdering av dyrevelferd: Qualitative Behaviour Assessment (QBA), som kan brukes for å vurdere og analysere adferd basert på kvalitative observasjoner. Denne metoden søker å gjøre subjekt-baserte vurderinger av dyr. Der kvantitative metoder observerer konkrete forhold ved dyrets adferd utenfra, forsøker QBA å innta dyrets perspektiv. Målet er å danne seg et bilde av "hvordan-det-er-å-være" dette individet. For å få til dette kreves ikke bare objektiv observasjon, men en anvendelse av vår egen empati i møte med et annet følelse vesen. Slik mener Wemelsfelder at vi kan fange opp informasjon som objekt-baserte metoder går glipp av, og ved å bruke både kvalitative og kvantitative metoder i en samlet vurdering kan vi lettere forstå hva det enkelte dyret trenger for å ha så god velferd som mulig. (Wemelsfelder, 1997)

Generelt sett er altså målet med kvalitative adferdsvurderinger å gi et rikere bilde av adferd enn bare kvantitative målinger alene kan. Dette gjøres med tanken om å gjøre det mulig å fange opp nyansene og konteksten i dyreadferd, og det kan derfor være spesielt verdifullt i forbindelse med dyrevelferdsvurderinger. QBA tilstreber å gi en mer dyptgående tilnærming til adferdsevaluering, men som med alle målemetoder som innebærer subjektive vurderinger er det viktig å være oppmerksom på utfordringene knyttet til subjektivitet og behovet for opplæring og kalibrering av observatører.

QBA-metoden innebærer å ikke bare evaluere hva dyr gjør, men også hvordan de gjør det. Observatører gjør subjektive vurderinger av adferden basert på deres intuitive opplevelse, kunnskap om artsspesifikk adferd, og opplæring i metoden. Det finnes to hovedtilnærminger til metoden: bruk av "free choice profiling" (FCP), eller en "fixed list" (FL). (Clarke et al., 2016)

"Free choice profiling" er en åpen tilnærming. Observatøren velger selv hvilke uttrykk man vil bruke for å beskrive adferden. Man er ikke begrenset av en fastsatt liste, og kan vurdere og beskrive dyrenes adferd med egne ord. Dette gir muligheten til å fange opp et bredere spekter av adferd og lar observatøren score akkurat det den oppfatter. Den er derimot vanskeligere å standardisere, da individuelle tolkninger kan variere. (Clarke et al., 2016)

Ved bruk av en "fixed list" får observatørene en liste over forhåndsdefinerte adferdsuttrykk man ser etter. Dette gjør det lettere å standardisere scoringene, og veileder også observatørene i hva de ser etter når de vurderer adferd (Clarke et al., 2016). Utfordringene ved bruk av FL kan være at man ikke får beskrevet adferden man observerer med sine egne ord, og man er nødt til å sette sine egne tanker inn i de forhåndsbestemte adferdsuttrykkene. Man trenger også en god forståelse for adferdsuttrykkene, som kan kreve en del diskusjon.

Ved å score observasjonsgraden av en rekke ulike emosjoner eller sinnstilstander, søker man å samle informasjon om dyrets opplevelse av sin situasjon i tidsrommet det ble observert (PAW, u.å.).

### **4.2.3 Reliabilitet**

For at en vurderingsmetode skal kunne brukes i stor skala, bør den ha god reliabilitet. God reliabilitet vil si at metoden gir samme resultat uavhengig av hvilken observatør som bruker den og på hvilket tidspunkt, så lenge det som vurderes er det samme. Ofte krever en vurderingsmetode at observatøren besitter visse forkunnskaper om temaet for at metoden skal være reliabel. Det er viktig å merke seg at det at en metode har reliabilitet, ikke sier noe om dens validitet - altså hvorvidt resultatene gjenspeiler virkelige forhold slik de faktisk er. (EFSA Panel on Animal Health and Welfare, 2012) Siden velferdsvurdering med QBA baserer seg på en i stor grad subjektiv vurdering av dyret, er det desto viktigere at metoden er reliabel. Med andre ord: man trenger å vite om ulike observatører, gjennom sin subjektive opplevelse, konsekvent fanger opp de samme aspektene ved dyrenes tilstand.

### **4.2.4 QBA i forskning**

Wemelsfelder tok til å begynne med utgangspunkt i Free Choice Profiling (FCP). I prosessen ble over 60 FCP-studier gjennomført - på blant annet gris, storfe, sau og høns, både på individ- og gruppenivå. De fant signifikant samsvar mellom observatørene, uavhengig av fagbakgrunn (Wemelsfelder, 1997).

I neste fase benyttet Wemelsfelder adferdsuttrykk som var blitt generert i FCP-studiene til å lage standardiserte lister over adferdsuttrykk til utprøving av det som nå er blitt QBA-verktøyet; en liste på 20 adferdsuttrykk med hver sin 125 mm visuelle analoge skala (VAS). Observatører scorer VAS basert på sin subjektive opplevelse av i hvor stor grad dyret viser dette adferdsuttrykket i løpet av observasjonen.

For å undersøke reliabilitet på ulike arter, har også mange andre forskere gjort studier av QBA. Med "Fixed List"-tilnærmingen har man blant annet sett på reliabilitet i forsøk på sau (Phythian et al., 2013), storfe (Andreasen et al., 2013) og hund (Stubsjøen et al., 2020). I alle disse tilfellene viste metoden generelt høyt samsvar, hvilket indikerer at metoden har god reliabilitet på disse artene.

#### 4.2.5 QBA på katt

Fra tidligere er kun 2 artikler publisert der forskere har studert bruken av QBA på tamkatten, *Felis catus*. I 2021 publiserte tidsskriftet *Animal Welfare* artikkelen *“Do you see the same cat I see? Relationships between Qualitative Behaviour Assessment and indicators traditionally used to assess temperament in domestic cats”* av I.C.Travnik og A.C. Sant’Anna i Brasil (Travnik & Sant’Anna, 2021). Artikkelen presenterer en studie der man sammenlignet QBA-scoring med standardiserte, kvantitative tester. Kattene ble filmet og utsatt for 4 tester: “Unfamiliar person” (UP), “Novel object” (NO), “Conspecific reaction” (CR) og “Food offering” (FO). Kattenes reaksjon på disse testene ble registrert ved hjelp av et etogram slik at resultatene kunne analyseres. Fremgangsmåten i testene er beskrevet i artikkelen.

Neste artikkel, *“Do you see the same cat that I see? Inter- and intra-observer reliability for Qualitative behaviour Assessment as temperament indicator in domestic cats”* ble publisert i samme tidsskrift i 2022 (Travnik et al., 2022). Her presenteres en studie i inter- og intra-observatør-reliabilitet av QBA-scoring mellom 19 observatører som hadde varierende grad av erfaring med katter, adferdsvurdering og QBA. Videomaterialet som ble vurdert var det samme som i den første studien. Analyse av dataene ble gjort med prinsipalkomponentanalyse, en metode som er utenfor omfanget av vår oppgave.

Disse studiene gir mye informasjon om metoden. For det første viste prinsipalkomponentene fra QBA-analysen høy korrelasjon (altså at resultatene sammenfaller) med “Unfamiliar person”- og “Food offering”-testene, og middels korrelasjon med de andre testene. Korrelasjon med kvantitative, objektive metoder bidrar til å validere QBA-metoden ved å gi den en forankring i et eksisterende rammeverk for vurdering av dyrevelferd. For det andre viste de ulike komponentene også høyt samsvar mellom alle observatører i reliabilitetstesten.

### 4.3 Bakgrunn for kattens adferd

For å kunne forstå kattens adferd og hvordan vi bør tilnærme oss den, er det nyttig å gjøre et tilbakeblikk på hvordan katten har utviklet seg og hvilke prosesser som har gjort seg gjeldende som utspring for de adferdsmønstre katten styres av i dag.

#### 4.3.1 Domestisering

Kattens samspill med mennesket ble tidligere antatt å ha begynt for rundt 3.600 år siden i oldtidens Egypt. Imidlertid tyder nyere arkeologiske funn på at domestiseringen startet betydelig tidligere, for nesten 10.000 år siden, i området kjent som “den fruktbare halvmåne”, som omfattet store deler av Midtøsten inkludert Nord-Egypt. Disse funnene oppsummeres i artikkelen *“The Taming of the Cat”* (Driscoll et al., 2009). Funnene er viktige, fordi de viser at kattens domestisering kom i gang i takt med det tidligste jordbruket, som regnes å ha startet i samme område rundt denne tidsperioden. Jordbruket førte med seg flere forhold som etter hvert ville bli gunstige for katten. For det første førte høstingen og lagringen av ulike typer mat til at husmusen (*Mus musculus*), flyttet inn i landsbyene, der de både fant ly og næringsgrunnlag. Villkatten, som i stor grad levde av små byttedyr som mus, forsøkte naturligvis å følge etter. Husmusen opplevdes som et problem for menneskene i disse landsbyene, som ville ha maten sin for seg selv. Dermed ble det dannet et gjensidig nytteforhold mellom katt og menneske som begge var tjent med.

Katten var dessuten godt utstyrt til å ta seg av giftige dyr som slanger, som ellers kunne skape problemer for menneskene. I tillegg til byttedyrene ble villkatten sannsynligvis tiltrukket av næringen den kunne få fra avfallshaugene utenfor landsbyene, ikke ulikt utefødte katter i vår tid. (Driscoll et al., 2009)

Man kan tenke seg at en annen fordel for kattene var at menneskene kunne forsvare landsbyene sine mot større rovdyr. I naturen hadde villkatten på grunn av sin størrelse en rolle som både jeger og byttedyr, mens fraværet av predatorer kan ha gjort det betydelig lettere å overleve, spesielt for hunnkatter med små unger.

Over tid, ettersom jordbruket spredte seg utover i verden, fulgte katten etter. Dens viktigste funksjon forble det samme: skadedyrkontroll. Akkurat når huskatten begynte å bli et kjæledyr som godtok direkte håndtering er uklart, men grunnet manglende kontroll på artens formering regner man med at det kan ha tatt svært lang tid. (Driscoll et al., 2009) Det noe indirekte forholdet mellom våre to arter, samt en relativt sett mer nylig domestisering (sammenlignet med f.eks. hund for over 15.000 år siden (Irving-Pease et al., 2019)), kan være med på å forklare hvorfor katten i dag ofte ses på som et selvstendig husdyr. Det kan også være med på å forklare hvorfor katter, spesielt de som har hatt liten menneskekontakt, kan vise stor varsomhet, frykt og til tider aggresjon i møte med mennesker. I tillegg er katter solitære jegere, mens hunden for eksempel er et flokkdyr med et kroppsspråk som er tilpasset et betydelig større sosialt samspill. Dette kan være årsaken til at mange mennesker i utgangspunktet opplever kattens kroppsspråk som vanskeligere å tolke.

#### **4.3.2 Avl**

Alle frittlevende arter har utviklet seg via evolusjon. Deres fysiske egenskaper og behov, deres instinkter og sosiale strukturer er først og fremst tilpasset det miljøet arten har utviklet seg i, og ikke nødvendigvis de omgivelser mennesker måtte velge å plassere dyrene i.

Gjennom målrettet avl har vi mennesker evnet å endre raser av mange arter til bedre å passe inn i menneskers dyrehold og bruksønsker. Derved har for eksempel hunderasene Weimaraner (stor jakt- og sporingshund) helt andre egenskaper, instinkter og behov enn en Bichon Havanaise (liten selskapshund) (Fogle, 2006)

Målrettet oppdrett av arter vi betegner som produksjonsdyr (i hovedsak arter som også anses å være flokk- eller gruppedyr), har også avstedkommet nye genetiske variasjoner mellom ulike raser. Noen av rasene var historisk ulike i utgangspunktet, mens andre raser har oppstått gjennom målrettet avl med kryssninger, som for eksempel pelssauen.

Opprinnelig ble tamkatten kategorisert som en underart av villkatter og ble da betegnet som *Felis silvestris catus*, mens den i dag som regel regnes som en egen art, *Felis catus*. Målrettet avl av tamkatten har også avstedkommet inndeling i ulike nye raser. Selektiv avl til det som nå betegnes som ulike katteraser startet rundt 1850 (Andersen & Halleraker, 2023; Braastad, 2019), og har i all hovedsak fokusert på utseende framfor andre egenskaper. Noen raser har dog adferdsmønster som skiller seg noe fra de andre rasene, som for eksempel Siamesere som generelt mye mer stemmebruk enn andre raser (Alderton, 1996; Braastad, 2019).

Blandingsraser, som går under fellesbetegnelsen huskatt, er de kattene det finnes langt på vei flest av i Norge. Selv om disse kattene stadig får innblanding av gener fra enkeltraser, har de formert seg fritt i så stor grad at eventuelle avlsrelaterte forskjeller i adferd er sjeldne å se. (Braastad, 2019)

### 4.3.3 Katteadferd i faglitteraturen

Den tyske zoologen og forskeren Paul Leyhausen (1916-1998) studerte katter over mange år og var forfatter av boka som fremdeles holder gullstandarden i tolkning og forståelse av tamkattens adferd og kroppsspråk. Første utgave, "*Das Verhalten der Katzen*", ble publisert i 1956 (Leyhausen, 1979). Gjennom årene kom den i stadig nye opplag (med noen revideringer) og disse ble oversatt til mange språk. Leyhausens observasjoner, illustrasjoner og tolkninger gjengis og refereres fremdeles i fagbøker om *Felis catus* (Braastad, 2019; Bradshaw et al., 2013) - Leyhausens studier kan derfor sies å ha dannet grunnlaget og formet hvordan vi i dag tolker adferd og kroppsspråk hos katt.

### 4.3.4 Shelterkatter

Dyrevernorganisasjoner ivaretar mange hjemløse katter i sheltere til de får nye, permanente hjem. Disse dyrene er hentet fra vanskelige situasjoner, ofte fra et liv alene utendørs. Katter på shelter kan ofte være traumatiserte, eller de kan ha hatt lite menneskekontakt og føle stor frykt i møte med mennesker. Vurdering av shelterkatters velferd kan være utfordrende, fordi disse kattene ofte vil være tilbakeholdne med å respondere på stimuli i vårt nærvær: alt vi klarer å observere er kattens frykt. Dette utgjør ikke bare et problem for vurdering av kattens emosjonelle tilstand, men også andre velferdsparametre som fysisk helse. Det utgjør dessuten en betydelig risiko å interagere med katter som, grunnet sin emosjonelle tilstand, er tilbøyelige til fysisk aggresjon. Desto viktigere er det at vi har gode, hensiktsmessige og trygge verktøy når vi forsøker å vurdere disse kattenes velferd. Slik kan vi også potensielt identifisere hvilke tiltak som må til for å hjelpe dem ut av en tilstand av frykt og stress.

## 4.4 Formål

Hovedmålet med studien er å bidra til å kvalitetssikre QBA som metode for å vurdere velferd hos katt. Herunder skal den:

Som delmål 1: undersøke samsvar mellom ulike observatører ved bruk av metoden (inter-observatør-reliabilitet)

Som delmål 2: undersøke effekten av kalibrering over tid på inter-observatør-reliabilitet

Som delmål 3: undersøke i hvilken grad hver observatør gjør konsekvente vurderinger over tid (intra-observatør-reliabilitet)

Studien utføres i samarbeid med Veterinærinstituttet, og de innsamlede dataene skal brukes videre i en vitenskapelig artikkel om samme tema.

## 5. Materiale og Metoder

### 5.1 Gruppeinndeling

Totalt deltok 6 dyrepleierstudenter i studien. Gruppe 1 bestod av de 3 forfatterne av oppgaven, som deltok i alle scoringer og kalibreringer. I gruppe 1 hadde alle studentene mye erfaring med katt og interesse for katteadferd, og hadde lest de grunnleggende artiklene om opphavet til QBA-metoden, samt artikler om metoden brukt på katt, på forhånd. (Travnik & Sant'Anna, 2021; Travnik et al., 2022; Wemelsfelder, 2007) Gruppe 2 bestod av 3 studenter fra samme årskull, som kun deltok i videoscoring 1 og videoscoring 2. I gruppe 2 hadde en av studentene nylig skaffet katt, mens de andre to kun hadde erfaring med katt fra klinikk, og ingen hadde kjennskap til metoden på forhånd.

### 5.2 Kvalitative adferdsuttrykk

For at QBA-metoden skulle gi mer konsekvente resultater blant de ulike observatørene, ble det i forkant av scoringene etablert en enighet om hvilke adferdsuttrykk man skulle se etter, og hvordan disse skulle defineres. Uttrykkene som ble valgt ut er basert på tidligere studier av QBA på katt (Travnik & Sant'Anna, 2021; Travnik et al., 2022). Veilederne oversatte uttrykkene fra engelsk til norsk, og gjorde noen endringer og omdefineringer der de mente det var hensiktsmessig. Listen over uttrykk og definisjoner følger i tabell 1.

**Tabell 1: Adferdsuttrykk og definisjoner**

Rolig (Calm)	I en fredelig mental tilstand, virker ikke urolig.
Ukomfortabel (Uncomfortable)	Katten viser fysisk eller mentalt ubehag i situasjonen, kan forsøke å unngå en kilde til ubehag, kan være nedstemt eller smertepreget.
Selvsikker (Confident)	Ikke avventende, virker sikker og trygg.
Fryktsom (Fearful)	Virker nervøs, lettskremt og urolig. Kan også virke overdrevent årvåken overfor omgivelsene.
Vennlig (Friendly)	Katten viser vennlighet mot andre katter og/eller mennesker.
Frustrert (Frustrated)	Uten mulighet til å tilfredsstille en motivasjon, kan virke irritert, anspent og/eller oppskaket.
Avslappet (Relaxed)	Virker ubekymret og uberørt, og ikke anspent.
Anspent (Tense)	Virker spent, bekymret og/eller engstelig.
Sosial (Sociable)	Tar initiativ til sosial interaksjon med andre katter og/eller mennesker, oppsøker andre katter.
Stresset (Stressed)	I en tilstand av anspenthet og angst, kan vise repeterende adferd.
Interessert (Interested)	Oppmerksom på og engasjert i miljøet rundt seg, men ikke på en oppskaket eller nervøs måte.

Avventende (Hesitant)	Motvillig, tilbaketrukket og årvåken.
Tilfreds (Content)	Virker fornøyd og opplever kontroll over sitt miljø.
Opphisset (Agitated)	Virker ansent, forstyrret og urolig, kan overreagere. Adferdsuttrykk kan være både motoriske og ikke-motoriske (f.eks. vokalisering).
Tillitsfull (Trustful)	Viser positiv interesse overfor andre katter og/eller mennesker, avslappet, forsøker ikke å gjemme seg. Kan være hengiven og søke oppmerksomhet.
Urolig (Uneasy)	Virker å ikke kunne slappe av, kan skifte posisjon ofte, eller gå urolig fram og tilbake.
Fredfull (Tranquil)	Virker veldig tilfreds, fredelig og uforstyrret, nyter øyeblikket.
Nøytral (Indifferent)	Er oppmerksom på omgivelsene men er ikke aktivt engasjert, kan virke likeglad og uinteressert.
Aktiv (Active)	Engasjert eller opptatt av noe, uavhengig av motorisk aktivitet, kan være i ro og samtidig være oppmerksom.
Undersøkende (Exploratory)	Uttrykker et ønske om å aktivt undersøke omgivelsene eller andre katter.

### 5.3 Visuell Analog Skala (VAS)

En visuell analog skala (VAS) er en skala som består av en rett linje med to endepunkter (minimum og maximum), men uten noen videre inndeling. Denne ugraderte skalaen er i mange år blitt brukt i det humane helsevesenet som et reliabelt og validert verktøy for egenrapportering av smertegrad (Begum & Hossain, 2022)(Sa42). VAS har i nyere tid også blitt et vanlig verktøy innen digital forskning (Reips & Funke, 2008). Med QBA-metoden scores dyrenes adferdsuttrykk på en VAS på 125 mm (Vedlegg 3), der hver observatør setter en strek på skalaen som representerer i hvilken grad observatøren mener at dyret uttrykker det spesifikke adferdsuttrykket, fra helt fraværende til dominerende gjennom observasjonsperioden.

### 5.4 Gjennomføring av studien

#### 5.4.1 Videoscoring

Alle 6 studenter utførte QBA-scoring av 12 kattevideoer første og siste dag av perioden for datainnsamling. Det var 18 uker mellom første og andre videoscoring. Videoene var tatt opp av personale ved FOD gården (Foreningen ved Omplassing av Dyr) og DOOA (Dyrebeskyttelsen Norge Oslo og Akershus), og kattene som ble filmet var omplassingskatter på midlertidig oppstalling hos dem. Videoene var redigert ned fra sin opprinnelige lengde til ca. 2 minutter av veilederne, med fokus på å fange opp en variasjon av adferdsuttrykk totalt sett i videomaterialet.

I forkant av første videoscoring presenterte veilederne de 20 QBA-uttrykkene med utdypende beskrivelser (Tabell 1). Alle fikk så anledning til å øve seg på scoring av 3 videoer som ikke ble inkludert i datasettet. Etter en del diskusjon ble et av uttrykkene endret – “oppmerksom” ble endret til “interessert” da dette opplevdes å presisere en positiv interesse.

Scoring ble så utført fortløpende og uten kommunikasjon mellom studenter eller veiledere. Hver video ble vist separat. Etter hver videovisning ble ett A4 ark med de 20 QBA-uttrykkene og tilhørende 125 mm Visuell Analog Skala (VAS) (Vedlegg 3) fylt ut av hver student for hver video/katt. Hver student hadde tilgjengelig et ark med de utdypende beskrivelser for hvert QBA-uttrykk for referanse.

Ved 2. videoscoring var det ingen presentasjon av, eller samtale rundt, adferdsuttrykkene. Samme videoer ble vist og scoret på samme måte, men videoene ble vist i randomisert rekkefølge. Ved ny videoscoring forventet man å finne en påvirkning av effekten observer-drift. Observer-drift defineres slik: “gradual, systematic changes over a period of time by a particular observer in their application of criteria for recording or scoring observations”. (APA, 2023)

#### **5.4.2 Kalibrering av observatørene**

Ett av delmålene med studien var å undersøke effekten av kalibrering over tid på samsvaret mellom observatørens scoring av adferdsuttrykk. Gruppe 1 gjennomførte 4 runder med kalibrering ved direkte observasjon av katter i månedene mellom første og andre videoscoring og før feltscoreingen.

2 sheltere i Oslo (FOD-gården på Klemetsrud og DOOA hjelpesenter på Mosseveien) ble besøkt 2 ganger hver, og studentene QBA-scoret mellom 3 og 6 katter per besøk. For at det skulle være praktisk gjennomførbart å score mange katter i løpet av en dag, ble det valgt en tidsbegrensning for hver observasjon på 2 minutter per katt. Kattene ble kun scoret basert på adferden som kom til uttrykk i løpet av disse 2 minuttene.

Kattenes oppholdstid på gården varierte stort. Kattene vi observerte levde i ulike miljø. Noen var i store rom, andre var i mindre rom, isolatbur eller tilvenningsbur der de hadde tilgang til skjul og en hylle å ligge på. Alle kattene hadde muligheten til å trekke seg unna og gjemme seg ved behov, og noen hadde tilgang til luftegård via katteluke. Miljøberikelsene de hadde varierte også stort mellom de ulike rommene de bodde i. I noen tilfeller hadde de tilgang til leker og andre innredninger. Noen hadde tilgang til å komme bort til oss, og andre så vi på gjennom en vegg. Som regel bodde de i grupper på to eller flere katter, men de isolerte bodde alene. I feltscoreingen prøvde vi å gjøre alle scoringene så like som mulig. Alle observatører sto samlet og så katten fra samme sted.

Mellom hver scoring sammenlignet og drøftet studentene sine scoringer for derved å fremme en felles forståelse og tolkning av både QBA-adferdsuttrykkene og kattenes adferd. Scoring fra disse shelterbesøkene ble ikke lagret eller inkludert i studiens datasett. Gruppe 2 gjorde ingen kalibrering, slik at effekten av kalibrering kunne tydeliggjøres ved å sammenligne de to gruppene i etterkant.



### 5.4.3 Feltscoreing

Feltscoreingen skjedde på FOD-gården uken før siste videoscoreing, på katter i den samme settingen som ved kalibrering. Observatørene fra gruppe 1 observerte også her hver katt i 2 minutter, fra et ståsted som så godt som mulig lot alle observatørene se kattens adferd fra samme synsvinkel. Scoreingen ble gjennomført uten kommunikasjon studentene imellom, og ble utført over 2 dager: 11 katter den første dagen, 9 katter den andre dagen.

## 5.5 Databehandling og statistikk

I løpet av studien ble det produsert 3 ulike datasett: disse kom fra henholdsvis videoscoreing 1, feltscoreing, og videoscoreing 2, i den rekkefølgen. Ett datasett inneholdt alle scoringer fra alle observatører i den gitte delen av studiet. Etter at hver scoring var ferdig, ble hver VAS målt med linjal fra venstre til høyre, og scoren ble ført inn i Microsoft Excel som en verdi mellom 0 og 125. Det ble godtatt en feilmargin på 1 mm. Målingen av alle dataene fra hver sesjon ble uavhengig utført av to av studentene i gruppe 1. Avvikene mellom de to målingene ble deretter kontrollmålt av den tredje studenten. Siden videoene i videoscoreing 2 kom i randomisert rekkefølge, måtte disse arrangeres i den opprinnelige rekkefølgen, slik at endringen over tid kunne analyseres per video.

### 5.5.1 Gjennomsnitt og variasjonsbredde

Dataene fra alle observatørene ble ført inn i Microsoft Excel, og gjennomsnitt og variasjonsbredde ble regnet ut for hver variabel (hvert adferdsuttrykk). Hensikten med dette var å undersøke i hvor stor grad hvert uttrykk ble observert. Dersom et uttrykk har lavt gjennomsnitt, betyr det at uttrykket observeres i liten grad, noe som kan være med på å forklare lav reliabilitet. Det samme gjelder uttrykk med lav variasjonsbredde (der kun en liten del av skalaen er brukt), spesielt i nederste del av skalaen.\*

### 5.5.2 Kendall's *W*

Inter- og intra-observatør-reliabilitet ble analysert av en av veilederne ved bruk av statistikkprogrammet Stata version MP 17.0 (Statacorp, College Station, Texas). Kendall's coefficient of concordance (Kendall's *W*) er et statistisk mål på samsvar mellom observatører. Resultatet for hver variabel blir et tall mellom 0 og 1, der 0 viser intet samsvar, mens 1 viser at alle observatørene var helt enige (Zaiontz, u.å.). Kategoriseringen av dataene er den samme som i tidligere studier av QBA (Travnik et al., 2022):  $W < 0,2$  regnes som veldig lavt samsvar,  $W = 0,2-0,39$  er lavt samsvar,  $W = 0,4-0,69$  er moderat samsvar,  $W = 0,7-0,89$  er høyt samsvar, og  $W > 0,89$  er veldig høyt samsvar. For å undersøke inter-observatør-reliabilitet ble Kendall's *W* for hver variabel fra videoscoreingene beregnet for alle observatører samlet, for gruppe 1 alene, og for gruppe 2 alene. Inter-observatør reliabilitet ble også undersøkt mellom observatørene i gruppe 1, basert på de direkte observasjonene i felt. Intra-observatør-reliabilitet ble undersøkt for hver enkelt observatør i begge gruppene mellom videoscoreing 1 og 2.

## 5.6 Litteratursøk

Istedenfor det tradisjonelle litteraturstudium, presenterer vår bacheloroppgave primærforskning i form av en pilotstudie der vi har samlet inn og analysert egne data. Metoden var allerede fastlagt, da vår studie og våre data er del av et større forskningsprosjekt om QBA som verktøy til å evaluere velferd hos katter.

Våre litteratursøk fokuserte derfor på å finne litteratur og artikler som

- Økte vår forståelse for hensikten med og bakgrunnen for QBA
- Beskrev tidligere QBA-forskning
- Beskrev og drøftet begrepene dyrevelferd og 'animal welfare'
- Beskrev andre metoder å evaluere dyrevelferd
- Beskrev og drøftet uttrykkene 'wellbeing', 'quality of life'
- Økte vår forståelse for analysemetodene som var blitt brukt i annen forskning
- Økte vår forståelse for analysemetodene som skulle brukes på våre data

Det ble søkt i Oria, PubMed og Medline

- Artikkeltitler og boktitler fra referanselister i allerede leste artikler
- Navn på forfattere som var hovedforfatter i relevante artikler allerede tilgjengelig
- Enkelt-termer på norsk og engelsk (som velferd, adferd, velvære) og akronymer (som QBA, PCA)
- Ordkombinasjoner som 'cats + welfare', 'animal + welfare', 'Qualitative Behavioural Assessment', 'animals + quality of life'

I tillegg benyttet vi oss av Google-scholar søk som ledet oss til informative og nyttige nettsider, blant annet i vår søken etter lett forståelige beskrivelser av analysemetodene Kendall's Coefficient of Concordance og Principal Component Analysis.

Utover dette er våre kilder både artikler gitt oss av våre veiledere, egne fagbøker og referanser til andre artikler.

## 6. Resultater

### 6.1 Gjennomsnitt og variasjonsbredde

Dersom et adferdsuttrykk brukes i veldig liten grad, eller kun bruker en liten del av VAS-skalaen, kan dette være med på å forklare lav reliabilitet. Tabell 2 viser gjennomsnitt og variasjonsbredde for alle uttrykkene i alle scoringsrundene.

Generelt brukte alle uttrykkene mesteparten av VAS-skalaen, uavhengig av scoringsrunde. De eneste uttrykkene som hadde noen nevneverdig lavere variasjonsbredde hadde det i feltscoringen, men selv her ble over 100 mm av skalaen brukt for alle uttrykkene. For videoscoringene var laveste gjennomsnitt 19,5 mm, mens laveste snitt i felt var for "oppisset" med 6,1 mm.

**Tabell 2: Gjennomsnitt og variasjonsbredde for alle tre scoringer**

Adferdsuttrykk	Videoscoring 1		Feltscoring		Videoscoring 2	
	Mean	Range	Mean	Range	Mean	Range
Rolig	69,4	0-125	68,6	0-125	68,7	0-125
Ukomfortabel	30,3	0-125	18,2	0-125	25,6	0-124
Selvsikker	76,8	0-125	65,1	0-125	74,0	0-125
Fryktsom	30,6	0-125	20,6	0-125	24,8	0-125
Vennlig	54,9	0-125	59,2	0-125	57,0	0-125
Frustrert	32,9	0-125	11,7	0-110	26,1	0-125
Avslappet	68,4	0-125	62,2	0-125	65,4	0-125
Anspent	38,6	0-125	21,1	0-125	30,5	0-125
Sosial	46,6	0-125	52,6	0-125	43,9	0-125
Stresset	34,6	0-125	9,6	0-103	24,0	0-125
Interessert	60,3	0-121	56,2	0-125	55,0	0-125
Avventende	29,8	0-125	27,7	0-125	19,5	0-125
Tilfreds	70,4	0-125	60,8	0-125	65,3	0-125
Oppisset	26,8	0-125	6,1	0-101	23,1	0-123
Tillitsfull	59,2	0-125	61,9	0-125	61,9	0-125
Urolig	45,7	0-125	17,2	0-125	39,7	0-125
Fredfull	60,4	0-125	44,0	0-125	58,3	0-125
Nøytral	31,2	0-115	23,4	0-125	34,6	0-125
Aktiv	65,4	0-125	57,3	0-125	75,3	0-125
Undersøkende	47,8	0-125	44,3	0-125	51,3	0-125

## 6.2 Delmål 1: Inter-observatør-reliabilitet

Resultatene for inter-observatør-reliabilitet er fremstilt i tabell 3 innenfor 4 kategorier, fra lavt samsvar til veldig høyt samsvar. Ingen av adferdsuttrykkene hadde veldig lavt samsvar ( $W < 0,2$ ) i noen av scoringsrundene, uavhengig av gruppesammensetningen. Generelt var samsvaret høyere for videoscoringene enn i felt, både før og etter kalibreringsperioden.

**Tabell 3: Inter-observatør-reliabilitet oppgitt i Kendall's  $W$**

Adferdsuttrykk	Videoscoring 1			Videoscoring 2			Felt-scoring
	Gruppe 1	Gruppe 2	Alle	Gruppe 1	Gruppe 2	Alle	Gruppe 1
Rolig	0,82	0,91	0,85	0,81	0,84	0,77	0,70
Ukomfortabel	0,75	0,80	0,74	0,59	0,83	0,64	0,65
Selvsikker	0,71	0,79	0,73	0,68	0,64	0,59	0,71
Fryktsom	0,70	0,81	0,71	0,57	0,73	0,58	0,57
Vennlig	0,81	0,86	0,82	0,84	0,78	0,77	0,90
Frustrert	0,68	0,78	0,65	0,68	0,66	0,58	0,61
Avslappet	0,84	0,82	0,82	0,81	0,79	0,74	0,70
Anspent	0,79	0,82	0,78	0,62	0,62	0,58	0,58
Sosial	0,83	0,82	0,80	0,77	0,88	0,81	0,87
Stresset	0,71	0,86	0,74	0,56	0,41	0,43	0,45
Interessert	0,63	0,45	0,46	0,81	0,60	0,63	0,84
Avventende	0,68	0,82	0,66	0,36	0,55	0,39	0,80
Tilfreds	0,87	0,96	0,87	0,84	0,81	0,79	0,71
Opphisset	0,70	0,82	0,70	0,61	0,76	0,59	0,27
Tillitsfull	0,80	0,82	0,78	0,79	0,78	0,73	0,80
Urolig	0,74	0,87	0,77	0,88	0,58	0,70	0,40
Fredfull	0,83	0,90	0,84	0,80	0,87	0,80	0,75
Nøytral	0,70	0,48	0,52	0,70	0,56	0,54	0,33
Aktiv	0,50	0,56	0,47	0,83	0,59	0,67	0,65
Undersøkende	0,82	0,76	0,75	0,85	0,82	0,76	0,85

Lavt	Moderat	Høyt	Veldig høyt
------	---------	------	-------------

### 6.2.1 Videoscoring

For videoscoring 1 havnet ingen av adferdsuttrykkene i kategorien **lavt samsvar** ( $W=0,2-0,39$ ). For videoscoring 2 var det ett uttrykk som hadde lavt samsvar mellom alle observatører og innad i gruppe 1 (avventende).

Fem uttrykk hadde **moderat samsvar** ( $W=0,4-0,69$ ) i videoscoring 1 mellom alle observatører. De fleste av disse hadde et lignende resultat innad i gruppene. For videoscoring 2 havnet 10 uttrykk i denne kategorien for alle observatører. Innad i gruppene havnet henholdsvis 7 og 9 uttrykk i denne kategorien for gruppe 1 og 2.

De fleste uttrykkene havnet i kategorien **høyt samsvar** ( $W=0,7-0,89$ ): 15 uttrykk blant alle observatører i videoscoring 1, og det var ingen betydelige avvik fra dette innad i gruppene. Ni av de samme uttrykkene hadde fremdeles høyt samsvar blant alle observatører i videoscoring 2. Her var samsvaret noe høyere innad i gruppene, henholdsvis 12 og 11 uttrykk her for gruppe 1 og 2.

For videoscoring 1 hadde 3 uttrykk **veldig høyt samsvar** ( $W>0,89$ ) innad i gruppe 2 (rolig, tilfreds og fredfull), mens ingen uttrykk havnet i denne kategorien blant alle observatører.

### 6.2.2 Feltscoring

I felt hadde 2 av uttrykkene **lavt** samsvar. 7 uttrykk hadde **moderat** samsvar, og 10 hadde **høyt** samsvar. Uttrykket "vennlig" var det eneste med **veldig høyt samsvar**.

### 6.2.3 Oppsummering

"Rolig", "vennlig", "avslappet", "sosial", "tilfreds", "tillitsfull", "fredfull" og "undersøkende" hadde konsekvent høyt samsvar i alle scoringer.

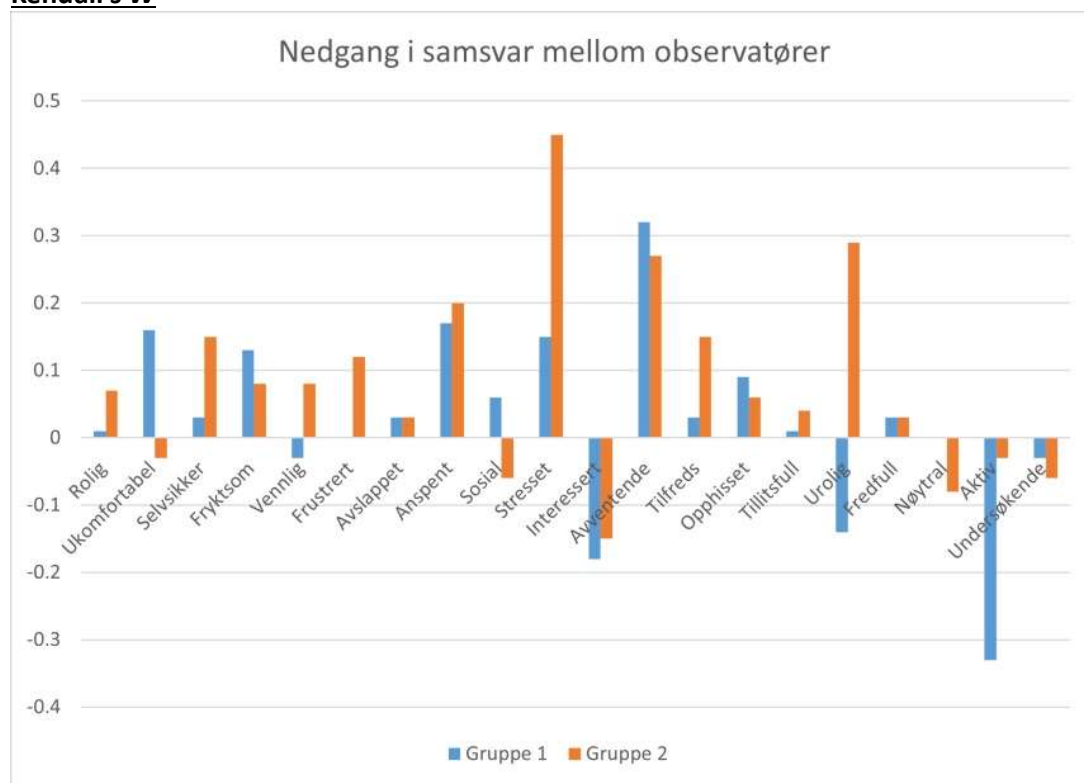
For videoscoringene lå resten av uttrykkene mellom moderat til høyt samsvar, med unntak av "avventende" som varierte betydelig.

I felt hadde, i tillegg til de ovennevnte, "selvsikker", "interessert" og "avventende" høyt samsvar. "Ukomfortabel", "fryktsom", "frustrert", "anspent", "interessert", "urolig" og "aktiv" hadde moderat samsvar. "Opphisset" og "nøytral" hadde lavt samsvar.

### 6.3 Delmål 2: Endring i inter-observatør-reliabilitet over tid

I denne studien ble de to videoscoringene utført med 4 måneders mellomrom. Man forventet derfor å finne en merkbar påvirkning av effekten “observer-drift” på samsvaret mellom observatører. Særlig gjaldt denne forventningen for samsvaret i gruppe 2, der fravær av noen form for kalibrering kunne bidratt til å øke denne effekten betydelig. Imidlertid gir resultatene ikke noe klart bilde av i hvilken grad endringen i samsvar skyldes observer-drift, effekt av kalibrering eller tilfeldigheter ved scoring. Noen uttrykk viste en nedgang i samsvar, mens andre viste en økning. Noen uttrykk endret seg mye, andre lite. Dette gjaldt i omtrent like stor grad for begge grupper. Nedgang i samsvar mellom videoscoringene vises som positiv akse i figur 1.

**Figur 1: nedgang i samsvar per gruppe mellom videoscoring 1 og videoscoring 2 oppgitt i Kendall's W**



### 6.4 Delmål 3: Intra-observatør-reliabilitet

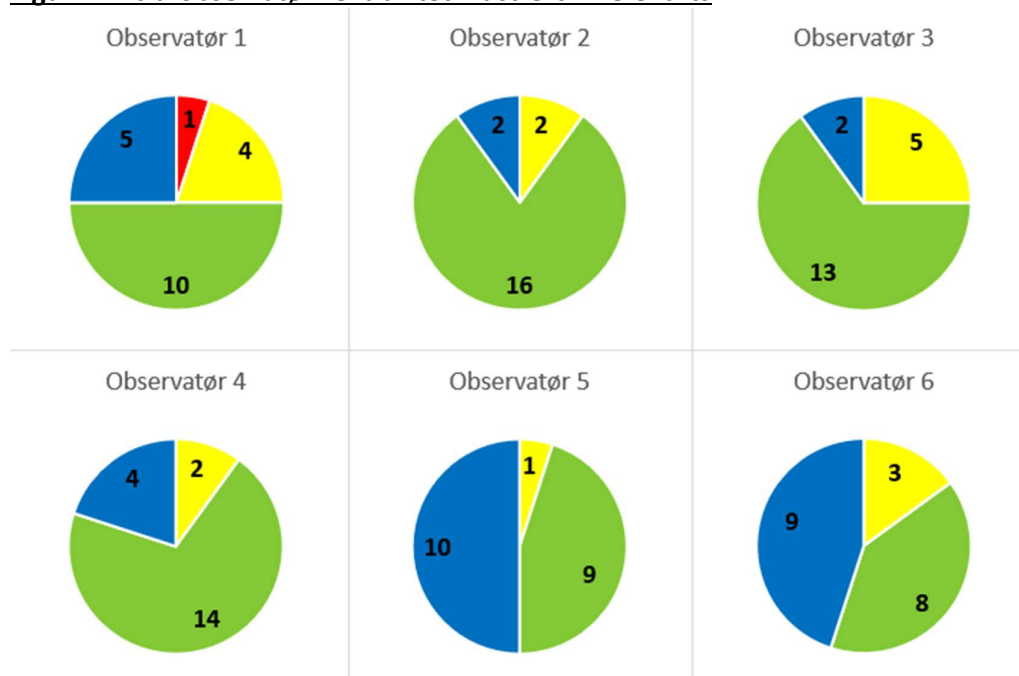
For å undersøke i hvilken grad hver observatør gjør konsekvente vurderinger over tid ved bruk av QBA, ble samsvar også analysert innad hos hver enkelt observatør før og etter kalibreringsperioden. Resultatene vises i tabell 4. Gruppe 1 består av observatør 1, 2 og 3, mens gruppe 2 består av observatør 4, 5 og 6. Generelt var intra-observatør-reliabilitet høy for alle uttrykk. De aller fleste havnet i kategorien høy, veldig høy, eller i øvre sjikt av moderat.

**Tabell 4: Intra-observatør-reliabilitet oppgitt i Kendall's W**

Adferdsuttrykk	Obs1	Obs2	Obs3	Obs4	Obs5	Obs6
Rolig	0,92	0,80	0,88	0,89	0,96	0,92
Ukomfortabel	0,74	0,75	0,63	0,85	0,90	0,78
Selvsikker	0,78	0,90	0,84	0,87	0,88	0,92
Fryktsom	0,66	0,66	0,60	0,71	0,80	0,87
Vennlig	0,92	0,88	0,82	0,95	0,84	0,88
Frustrert	0,82	0,77	0,83	0,73	0,94	0,61
Avslappet	0,90	0,86	0,88	0,84	0,81	0,93
Anspent	0,67	0,83	0,72	0,78	0,94	0,95
Sosial	0,80	0,83	0,79	0,95	0,90	0,83
Stresset	0,75	0,63	0,65	0,61	0,86	0,73
Interessert	0,87	0,89	0,74	0,74	0,65	0,69
Avventende	0,36	0,86	0,49	0,56	0,89	0,81
Tilfreds	0,85	0,84	0,92	0,89	0,95	0,93
Opphisset	0,75	0,73	0,72	0,83	0,86	0,77
Tillitsfull	0,92	0,85	0,79	0,91	0,92	0,92
Urolig	0,93	0,83	0,89	0,75	0,92	0,94
Fredfull	0,88	0,80	0,90	0,95	0,90	0,95
Nøytral	0,69	0,87	0,73	0,71	0,73	0,75
Aktiv	0,89	0,82	0,54	0,82	0,87	0,48
Undersøkende	0,68	0,92	0,85	0,83	0,92	0,90



**Figur 2: Intra-observatør-reliabilitet illustrert i Pie Charts**



## 7. Drøfting

### 7.1 Drøfting av funn

#### 7.1.1 Gjennomsnitt og variasjonsbredde

Lav gjennomsnittlig scoring av et adferdsuttrykk kan ha ulike potensielle årsaker. Det kan for eksempel hende at uttrykket er lite treffende for adferden til arten man ser på, eller at utvalget av katter som ble vurdert var for lite eller hadde for lite variert adferd til å plukke opp et spenn som man i teorien ville forvente å finne i en annen populasjon. Enkelte av uttrykkene bør dessuten antas å ha en lavere forekomst, avhengig av situasjonen. For eksempel fremstår “nøytral” som en lite sannsynlig respons fra katter på shelter som vi har sett på i vår studie, da dette fordrer at katten viser liten reaksjon på omgivelser og stimuli. Det er nærliggende å forvente at katter på shelter, spesielt de med kort oppholdstid, vil være oppmerksomme på det meste som skjer rundt seg. Sterkt negativt ladede uttrykk som “stresset” vil man ikke forvente å observere i stor grad i private kattehold, mens forekomsten av disse på et shelter vil variere svært mye avhengig av kattene som bor der.

Hvis et uttrykk har lav variasjonsbredde, betyr det at en liten del av skalaen er brukt totalt sett. I så fall vil hver mm gi større utslag ved utregning av Kendalls  $W$ , og dermed gir dataene mindre informasjon samtidig som at reliabiliteten blir lavere. I denne studien hadde ingen uttrykk lavere variasjonsbredde enn 100 mm, noe som er meget høyt. De uttrykkene hvor variasjonsbredden var relativt sett lavest, spesielt i felt, var de mest negativt ladede som “stresset” og “opphisset”. Årsaken til dette kan være at flertallet av kattene vi scoret var blitt godt vant til å være på shelteret. Selv om et øyeblikksbilde gir mindre utslag for disse uttrykkene, spesielt for katter med lang oppholdstid, fremstår det tydelig for oss at de treffer godt på enkelte katter, og de antas å ville treffe særdeles godt på katter som for eksempel er utefødte og har hatt liten menneskekontakt.

#### 7.1.2 Delmål 1: Inter-observatør-reliabilitet

Delmål 1 i denne studien var å undersøke inter-observatør-reliabilitet ved bruk av QBA-metoden på katt. Det ble brukt en Fixed List-tilnærming, med 20 predefinerte adferdsuttrykk.

Metoden viste generelt høy inter-observatør-reliabilitet for de aller fleste adferdsuttrykkene. Uttrykkene med konsekvent høyest reliabilitet var generelt de som beskriver positivt ladede sinnsstemninger. Dette gjaldt spesielt for “rolig”, “vennlig”, “avslappet”, “sosial”, “tilfreds”, “tillitsfull”, “fredfull” og “undersøkende”, som ga høyt-til-veldig-høyt samsvar i alle scoringer.

Samsvaret innad i gruppe 2 var høyere i snitt enn i gruppe 1, spesielt i videoscoring 2. Årsaken til dette kan ha vært at observatørene i gruppe 2, som jevnt over hadde lite erfaring med arten, dermed hadde et erfaringsgrunnlag som var mer likt enn i gruppe 1. At metoden gir høy reliabilitet for gruppen med lite erfaring indikerer både at metoden er intuitiv, at den ikke nødvendigvis krever mye erfaring med katt for å være reliabel, og at den første opplæringen i bruk av den fungerte godt. Det må riktignok nevnes at alle observatørene var dyrepleierstudenter som hadde hatt noe undervisning i katteadferd tidligere.



Inter-observatør-reliabiliteten til QBA-metoden har blitt undersøkt i diverse studier på ulike arter. Generelt har metoden vist høy reliabilitet, både med FCP- og FL-tilnærminger. I én studie (Clarke et al., 2016) ble de to tilnærmingene sammenlignet ved at de ble brukt av to ulike grupper for å vurdere det samme videomaterialet av griser. Konklusjonen i dette tilfellet var at både FCP og FL viste høy inter-observatør-reliabilitet, og at det ikke fantes noen signifikant forskjell mellom de to tilnærmingene. Videre har FL-tilnærmingen blant annet vist høy reliabilitet i studier på sau (Phythian et al., 2013) og storfe (Andreasen et al., 2013), og i en studie på hund fra 2020 (Stubsjøen et al., 2020) viste den høy reliabilitet, både generelt og for de fleste uttrykkene enkeltvis.

### **7.1.3 Delmål 2: Observer-drift og effekt av feltkalibrering**

Delmål 2 i studien var å undersøke effekten av kalibrering over tid på metodens inter-observatør-reliabilitet. Generelt var inter-observatør-reliabiliteten lavere i videoscoring 2 enn i videoscoring 1. Det er imidlertid ikke bare feltkalibreringen som kan ha ført til endring i samsvar over tid. En annen årsak til dette kan ha vært observer-drift. Begge grupper falt i inter-observatør-reliabilitet mellom videoscoring 1 og videoscoring 2 for de fleste uttrykk. At fallet i samsvar for de fleste uttrykkene var relativt likt hos gruppe 1 som hos gruppe 2 tyder på at kalibreringsprosessen i felt ikke hadde en signifikant effekt på disse uttrykkene. For å motvirke observer-drift kan det altså være viktig å utføre jevnlig kalibrering sammen med mer erfarne brukere av metoden, for eksempel ved bruk av videoklipp.

En annen årsak til fallet i samsvar hos gruppe 1 kan ha vært at hyppig bruk av metoden gjorde at vi begynte å score på automatikk: altså at vi brukte kortere tenketid på hver video i videoscoring 2 enn i videoscoring 1 og dermed scoret dem forskjellig. Det er i så fall viktig å passe på ved langvarig bruk av metoden at brukeren ikke blir blasert for den, da dette kan gi mindre konsekvente vurderinger. Riktignok er dette trolig et større problem der man scorer de samme videoene eller individene flere ganger.

### **7.1.4 Delmål 3: Intra-observatør-reliabilitet**

Delmål 3 i studien var å undersøke intra-observatør-reliabilitet ved bruk av metoden. Resultatene viser at intra-observatør-reliabiliteten var konsekvent høy. De eneste uttrykkene med et noe lavere resultat hos flere observatører var "avventende", "fryktsom" og "stresset", men også her var samsvaret høyt hos halvparten av observatørene. Generelt betyr dette at endringen i scoring av videoene for den enkelte observatør var liten. At dette var tilfellet for alle observatører indikerer at metoden gir relativt konsekvente svar over tid for den enkelte observatør (i hvert fall innenfor en periode på 4 måneder), både med og uten kalibrering. Dersom intra-observatør-samsvar hadde vært lavt for gruppe 1, ville dette vært en indikasjon på større effekt av kalibreringen på hvordan disse observatørene scoret hver katt.

Travnik (Travnik et al., 2022) pekte på at lengden på tidsperioden mellom scoringsrunder må antas å ha signifikant betydning for intra-observatør-reliabiliteten - jo lengre tid mellom scoring av de samme videoene, jo mindre sannsynlig at observatørene husker eller påvirkes av hvordan de scoret forrige gang. I studien deres hadde man et opphold på 10 måneder mellom scoringene, for å redusere risikoen for at observatørene skulle huske innholdet i de enkelte videoene.

I vår studie gikk det 4 måneder mellom de 2 videoscoringene. I tillegg ble rekkefølgen på videoklippene randomisert, av samme årsak. Vår opplevelse av videoscoring 2 var at vi ikke hadde noe klart minne av klippene eller hvordan vi tidligere hadde scoret dem. Vi tror at denne oppholdstiden var lang nok til at hukommelsen vår ikke påvirket intra-observatør-reliabiliteten.

## 7.2 Sammenligning med tidligere studier

To studier av QBA har blitt utført på katt. (Travnik & Sant'Anna, 2021; Travnik et al., 2022) Begge studiene brukte det samme videomaterialet. Videoene som ble vurdert ved bruk av QBA var på 12 minutter hver og sammensatt av enkeltvideoer av 4 kvantitative tester: "Unfamiliar person" (UP), "Novel object" (NO), "Conspecific reaction" (CR) og "Food offering" (FO). QBA-scoringen i den første artikkelen ble utført av kun 1 person, mens de samme videoene ble scoret av 19 observatører i studien som undersøkte reliabilitet. Av disse deltok 13 i neste scoringsrunde 10 måneder senere. I likhet med vår studie scoret man kattene på 20 ulike adjektiv/adferdsuttrykk, men de baserte reliabilitetsanalysene på hovedkomponenter fra prinsipalkomponentanalyse og resultatene kan derfor ikke sammenlignes direkte. Dette gjelder også for mange andre studier av QBA fordi man sjelden har analysert samsvar for enkeltuttrykk.

I studien av QBA på hund (Stubsjøen et al., 2020) analyserte man inter-observatør-reliabilitet både for prinsipalkomponentene og enkeltvis for adferdsuttrykkene. I hovedstudien scoret 22 veterinær- og dyrepleierstudenter 12 videoklipp av hunder på shelter. Som en del av en bacheloroppgave som inngikk i det samme prosjektet ble metoden også anvendt på hunder i felt. (Johannessen et al., 2020) Resultatene fra denne studien er dermed lettere å sammenligne med våre, selv om en del av uttrykkene som ble brukt er forskjellige. Ved analyse av samsvar i hovedstudien fant man, i likhet med vår studie, moderat-til-høyt samsvar for de fleste uttrykk. De eneste uttrykkene som hadde lavt samsvar, var "aggressive" og "calming" (konfliktdempende).

Studentene som brukte metoden på hund i felt fikk lavt samsvar for uttrykkene "ukomfortabel", "nedstemt", "kjeder seg", "nervøs", "aggressiv", "frustrert", "nøytral" og "stresset". Med unntak av "nøytral", var alle uttrykkene med lavt samsvar negativt ladede. Dette stemmer overens med vår erfaring: negative sinnsstemninger var vanskeligere å score, særlig i felt, mens positive sinnsstemninger var lettere både i felt og på video. Samtidig var det kun to uttrykk ("opphevet" og "nøytral") som fikk lavt samsvar i felt i vår studie. Det var heller ingen av uttrykkene som hadde lav variasjonsbredde, mens dette gjaldt i stor grad for de negativt ladede uttrykkene brukt på hund. At resten av våre negativt ladede uttrykk ikke hadde lavere samsvar, kan altså skyldes at de var mer treffende for dyrene i den situasjonen vi observerte dem, både på grunn av settingen og forskjeller i artens adferdsmønster.

## **7.3 Feilkilder**

### **7.3.1 Målinger utført for hånd**

Målingen av hver enkelt score ble gjort med linjal. I hver av videoscoringene var det med 6 observatører som scoret 12 katter på 20 uttrykk: dette gir 1440 målinger per runde. I felt var det med 3 observatører som scoret 20 katter på 20 uttrykk: dette gir 1200 målinger.

Så mange målinger utført for hånd gir potensielt en stor feilkilde ved innføringen av dataene som skal analyseres. Det krever høy konsentrasjon av den som måler, er en repetitiv oppgave, og tar mye tid. I vår oppgave ble målingene uavhengig utført to ganger, og deretter kontrollmålt ved avvik. Dermed ble denne feilkilden minimal, bortsett fra avvik på 1mm som vi ser på som akseptabelt da disse ikke vil ha noen signifikant effekt på resultatet.

### **7.3.2 Kalibreringsprosessen**

Veilederne på Veterinærinstituttet hadde i utgangspunktet gjort en avtale med DOOA om filming av videoer og besøk på hjelpesenteret deres. Ved første møte kom det imidlertid frem at DOOA kun hadde en håndfull katter inne, og ikke kunne si noe sikkert om inntak fremover. Dette var et for lite studieutvalg, spesielt for feltscoringen som, av hensyn til de senere analysene, helst skulle utføres på minst 20 individer. Siden Palmgren hadde kontakter i FOD, tok han tidlig kontakt med dem for å etterspørre hjelp med videoer og muligheten for kalibrering der. Uansett ble de to første kalibreringsrundene utført hos DOOA, med kun 3 scoringer hver. De påfølgende to kalibreringsrundene ble utført hos FOD, som hadde betydelig flere katter inne. Gjennomgang av metoden sammen med veilederne skjedde kun i forkant av videoscoring 1, der vi så de 12 videoene og fikk forklart hvordan vi skulle score de forskjellige adferdsuttrykkene. Videre kalibrering utførte vi utelukkende på egenhånd der vi diskuterte uttrykkene oss imellom.

Våre diskusjoner, spesielt angående uttrykk vi anså som vanskeligere å forstå, kan ha forvirret oss mer enn det hjalp - vår tolkning av uttrykkene kom mer i fokus, og vi videreutviklet vår egen forståelse innad i gruppen mer enn vi kalibrerte mot en forhåndsdefinert forståelse. Med andre ord: kalibreringen kan ha redusert observer-drift mellom oss som gruppe, uten at den i noen betydelig grad påvirket drift vekk fra metoden slik den ble presentert. Dersom denne effekten i realiteten var stor, ville man forvente å finne høyere samsvar innad i gruppe 1 ved videoscoring 2, og lavere intra-observatør-reliabilitet hos hver enkelt av observatørene. Imidlertid støtter ikke resultatene den opplevelsen vi hadde under kalibreringen - hver observatør scoret relativt konsekvent i videoscoringene, og samsvaret gikk ned i videoscoring 2 for de fleste uttrykk.

### **7.3.3 Mangel på variert adferd**

Katten er en art som tilbringer mesteparten av livet sitt i ro. (Sordo et al., 2020) Dette fikk vi oppleve da vi skulle teste verktøyet på katter i felt. Ved våre besøk hos DOOA og FOD var det vanskelig å finne katter som gjorde noe særlig annet enn å ligge eller sitte stille. De 12 videoene vi brukte til videoscoringene var derimot filmet over en lengre tidsperiode av personale ved shelterne, med til hensikt å fange opp ulike typer adferd. Videoene var så blitt redigert ned til 2 minutter. Videoene viste derfor betydelig mer variert adferd enn vi fikk se under våre besøk hos shelterne.

Kattene vi møtte på DOOA lå for det meste i ro, og viste lite variasjon i adferd. Selv om det var mulig å se på samme katt flere ganger i en øvingssammenheng for å fange opp ulik adferd når kattene var i bevegelse, fikk vi ikke sett så mye som vi håpet på. Det var enkelte uttrykk vi så svært lite tegn til. Dette ga lite rom for å diskutere hva vi så satt opp mot uttrykkene vi fulgte, da vi følte visse uttrykk krevde mer diskusjon for effektiv kalibrering.

Noe annet vi opplevde under kalibreringen var at kattene som lå og tilsynelatende sov ga størst samsvar, men de som lå stille men var på vakt var de som ga størst sprik i scoringene våre. Hvorvidt en katt som lå stille var rolig eller på vakt var til tider vanskelig å tyde ettersom små nyanser utgjorde stor forskjell i vår oppfatning.

Det utgjør en betydelig utfordring for felttesting av metoden at organisasjoner som driver med omplassering av katt som regel er frivillig drevet og har et inntaksmønster som sjelden er forutsigbart. Imidlertid er det positivt at samsvaret generelt var høyt etter første kalibrering på eksempelvideoer, hvilket bidrar til å gjøre mangelen på sammenlignbare studieobjekter til en mindre kritisk feilkilde.

### **7.3.4 Effekt av kattens miljø på observatøren**

I følge Wemelsfelder vil ikke miljøet katten befinner seg i påvirke observatørens oppfattelse i noen særlig grad for huskatter (Wemelsfelder, 2007), men under kalibrering opplevde vi at forskjellen i levemiljøet kunne i noen tilfeller gi utslag i scoringene for oss. Spesielt tydelig var dette for Enger og Wilberg som hadde lite tidligere erfaring i shelter. Vi så for eksempel at vi hadde lettere for å score de isolerte kattene høyere i uttrykk som “ukomfortabel” og “frustrert”, uten at katten nødvendigvis utførte denne adferden i større grad enn de som bodde mer fritt. Palmgrens tidligere arbeid i shelter kan ha gitt han en annerledes forståelse av disse. Forutinntatthet om at dyr i bur nødvendigvis innehar en mer negativ sinnsstemning er et eksempel på at eksterne forhold kan påvirke våre vurderinger av dyr. Et annet eksempel på dette ved bruk av QBA fant Tuyttens et al. (Tuyttens et al., 2014), som konkluderte med at observatører scorete høner annerledes dersom de ble fortalt at hønene var fra en økologisk gård sammenlignet med en konvensjonell gård. For kattene som var oppstallet sammen i større arealer opplevdes miljøet generelt å ha en mindre påvirkning på vår vurdering.

### **7.3.5 Tidsbegrensning: 2 minutter per katt**

Under kalibrering og under feltscoringen ble det satt en tidsramme for observasjon på 2 minutter per katt. Vi valgte å gjøre dette både for at scoringene skulle kunne utføres innenfor et overkommelig tidsrom, og for at prosessen skulle ligne mest mulig på videoscoring. Problemet med denne løsningen er at det ofte kan være vanskelig å fange opp helheten i en katts adferd på så kort tid. Vi opplevde flere ganger at adferden som katten utførte før eller etter de 2 minuttene var ulik den adferden vi hadde scoret, og i mange tilfeller ville gitt et bedre bilde av kattens helhetlige sinnsstemning.

I Welfare Quality protokollene, ved velferdsvurdering av produksjonsdyr der man ser på hele besetninger, er anbefalt observasjonstid med QBA opptil 20 minutter. (Welfare Quality Network, u.å.) Det synes ikke å finnes beskrivelser i tidligere studier av hvor lang observasjonstid som er nødvendig når man bruker metoden i velferdsvurderinger på enkeltindivider i felt. Travnik et al. (Travnik & Sant'Anna, 2021; Travnik et al., 2022) benyttet seg av 12 minutter lange redigerte videoopptak i sin QBA scoring av katter, men her var man også interessert i sammenhengen mellom ulike metoder.

Shelterkattene i vår studie ble enten:

1. Filmet i sin vanlige oppstallingslokalitet uten andre stimuli enn å bli filmet av en person de var vant til å omgås (videoscoring 1 og videoscoring 2).
2. Observert i 2 minutter av 3 observatører som hadde besøkt og observert dem 2 ganger tidligere under observatørens kalibrering og som derfor ikke var helt fremmede lenger (feltscoring)

Siden vi i denne studien i hovedsak undersøkte reliabilitet, er ikke tidsbegrensningen i seg selv noe stort problem, men vi erfarer at det ved en velferdsvurdering vil være nødvendig å bruke betydelig mer tid per katt enn vi gjorde her.

### **7.3.6 Eksponering for stimuli versus ren observasjon**

I tidligere QBA-analyse på katt utført av Travnik et al. (Travnik & Sant'Anna, 2021; Travnik et al., 2022) har kattene i tillegg blitt utsatt for ulike tester satt opp i en innhegning for å teste deres reaksjon i situasjoner den kan møte i en husholdning. Den første av testene var at en person kan ikke hadde kjennskap til skulle gå rolig inn i rommet, sette seg noen meter unna katten og rekke ut en hånd mot den. Denne testen kalles "Unfamiliar Person". Kattens fysiske reaksjon til en ukjent person fylles kvantitativt inn i et skjema, hvor de ser på blant annet haleposisjon, ørestilling og distanse katten rømmer fra observatøren.

I vår studie skulle vi ikke interagere med kattene. Kattene ble ikke utsatt for stimuli annet enn observatørens tilstedeværelse, for å se om ren observasjon uten direkte kontakt med mennesker også kan få fram pålitelige data på en gjennomførbar måte. Dette ville særlig være en fordel i tilfeller der kattens aggresjon ellers ville utgjøre en betydelig risiko for den som skal score.

Vi utførte altså ingen slike tester i vår studie, men likevel ble den samme påvirkningen som i "Unfamiliar Person"-testen relevant for oss under feltscoring.

Kattenes tilnærming til observatøren var en faktor vi ikke kunne påvirke. Selv om vi ikke hadde noen intensjoner om å utsette kattene for stimuli var dette ikke til å unngå i felt, da bare vår tilstedeværelse kan anses som en stimulus. Ved våre første møter med kattene på FOD og DOOA under kalibrering var det tydelig at ikke alle kattene var like begeistret for å se oss. De kunne løpe unna, gjemme seg eller sitte veldig stille og følge med på oss. Etter flere besøk hos de samme kattene var det klart at samtlige møter med observasjon og litt kos hadde en betryggende effekt. De viste større tegn til vennlighet mot en observatør de kjente fra før. Det virker sannsynlig at dette gjorde at de positivt ladede uttrykkene kom tydeligere frem og gjorde det mindre sannsynlig at kattene viste tegn til negativt ladede uttrykk. Vår kalibrering påvirket adferden til kattene vi senere scoret i felt, og dermed dataene vi samlet inn. For en reliabilitetsvurdering av verktøyet vil dette ha mindre å si, men kan være viktig å ta i betraktning ved en reell velferdsvurdering.

Det er tidsbesparende å ikke utføre noen konkrete tester, og at det gjennomføres uten direkte kontakt kan være en fordel når man ser på katter i et shelter som ikke alltid er vennlige mot mennesker. Ofte har man mange katter å se på, som i mange tilfeller heller ikke er så komfortable med menneskelig kontakt. Selv om man kanskje ikke får en like grundig vurdering som i studien til Travnik et al. vil man her ikke bruke mer enn noen minutter per katt (+ dataanalyse), hvor i deres studie kunne det ta flere uker. Observatøren trenger heller ikke interagere med kattene, som er i en tilvenningsfase til mennesker. Resultatene funnet uten bruk av stimuli ga fortsatt et høyt inter-observatør- og intra-observatør-samsvar. Dette indikerer at QBA-scoring kan gjennomføres ved bare en kort observasjon av katten. Vi vet ikke hvorvidt intervensjoner, å utsette kattene for ulike stimuli eller hendelser, ville endret våre scoringer og evt. inter- og intra-observatør-reliabiliteten.

### **7.3.7 Minimum og maximum**

Hvor mye av et uttrykk må katten vise for å score maksimalt i et uttrykk? Observatører med ulik erfaring med katt vil ofte ha forskjellig oppfatning av hvor sterkt en adferd potensielt kan komme til uttrykk. Vi var nødt til å foreta egne vurderinger for hva vi mente ga en full score. Selv om dette for det meste var greit, viste det seg til tider å være en utfordring. Spesielt kom dette til syne når det gjaldt uttrykk vi hadde begrenset erfaring med og med mer uklare definisjoner, for eksempel "stresset". Dette uttrykket er tenkt å beskrive en sterk negativ opplevelse for katten, som vi regner med at mange potensielle brukere av metoden har liten erfaring med i praksis. Uten en solid bakgrunn for hvor høyt vi skulle sette scoren, var det vanskelig å komme frem til en vurdering.

Scoring av minimum var lettere, men det er viktig å ta i betraktning at uttrykkene er individuelle og ikke baseres på hverandre i større grad enn det adferden til katten viser til. I enkelte tilfeller viste også kattene tydelige tegn til visse adferdsuttrykk som kunne vare bare i noen sekunder. Disse kortvarige uttrykkene kunne være spesielt utfordrende å sette en konkret score på, da de bare utgjorde et lite øyeblikk, men fortsatt kunne oppfattes veldig tydelig i dette korte tidsrommet.

### **7.3.8 Tap av fokus under scoring**

Under videoscoring 2 ble alle 12 videoklipp vist og scoret fortløpende, uten pause. 1 observatør rapporterte om slitenhet, nedsatt konsentrasjon og opplevelse av sammenblanding under de siste 4 videoene. Tap av fokus opplevdes også som et potensielt problem under feltscoringen. Den første dagen scoret vi totalt 11 katter, i diverse forskjellige bygg. Etter dette opplevde samtlige observatører betydelig nedsatt konsentrasjonsevne. Travnik et al (Travnik et al., 2022) innførte 30 minutter pause etter hver 4. videovurdering (12 minutter hver), samt instruks om å avbryte dersom man følte seg sliten eller uoppmerksom. Dersom man skal gjøre en velferdsvurdering og må bruke mer tid per katt, kan det altså være viktig å begrense antall dyr man vurderer i ett strekk.

## **7.4 Diskusjon av adferdsuttrykk**

### **7.4.1 Motpoler**

Scoringene vi gjorde med QBA skulle i utgangspunktet kun reflektere i hvilken grad man oppfatter det gjeldene uttrykket. Noen adferdsuttrykk kan sees på som motpoler, men fravær av et uttrykk betyr ikke at det motsatte må være til stede. En uvennlig katt scorer ikke nødvendigvis 0 i “vennlig” hvis den også viser tegn til vennlighet. Det er altså ikke gitt ut ifra metoden at noen av uttrykkene utelukker hverandre. Det vil derfor være en fordel om adferdsuttrykk som er motpoler fjernes fra utdypningene. Dette gjelder for utdypningene av “rolig”, “selvsikker” og “avslappet”.

### **7.4.2 Sosiabilitet**

Uttrykkene “vennlig”, “sosial” og “tillitsfull” handler om sosiabilitet. I definisjonen av disse finnes det ikke noe skille mellom sosiabilitet mot mennesker og sosiabilitet mot andre katter (eller andre dyr). Årsaken er at det ved analyse vil være umulig å skille mellom tilfeller der katten scorer lavt fordi den ikke er sosial, og tilfeller der den ikke har muligheten til å være sosial (ved fravær av henholdsvis andre katter eller mennesker). Vi ser likevel i praksis et stort behov for å skille mellom sosiabilitet mot mennesker og sosiabilitet mot katter, dersom metoden skal brukes til dyrevelferdsvurderinger eller for å se på adopterbarhet.

Omplasseringsinstitusjoner er helt avhengige av å vurdere hvordan katten vil reagere på sine nye eiere, og hvorvidt den passer i hjem med andre katter.

Sosiabilitet mot mennesker er også den faktoren som er vist å korrelere best med høy adopterbarhet. (Gourkow & Fraser, 2006) Dersom dette skillet ikke lar seg implementere i QBA, vil det være hensiktsmessig å bruke kvantitative tester som undersøker sosiabilitet i tillegg.

### **7.4.3 utfordringer knyttet til “Fixed List”**

#### **Definisjoner av uttrykkene**

Definisjonene hadde til hensikt både å utdype og presisere hva det enkelte adferdsuttrykket skulle innebære av tolkning eller opplevelse hos observatør, men vi opplevde at de i noen tilfeller kan ha forvirret oss mer enn de har hjulpet. I flere tilfeller virket det for oss som om definisjonene kompliserte uttrykkene og gjorde scoring vanskeligere. Dermed kan scoringen kanskje ha blitt mindre presis enn om vi hadde brukt vår egen forståelse. Under scoring opplevde vi stadig et (muligens unødig) behov for kontroll-lesing av definisjonene mens vi utførte scoring av hver enkelt katt. Dette kan være en utfordring ved bruk av en “fixed list”, som man ikke støter på med “free choice” tilnærmingen. Likevel viser det generelt høye samsvaret at dette opplevde problemet ga mindre utslag enn vi hadde inntrykk av der og da.

### **Adferdsuttrykk på to språk**

I listen over adferdsuttrykk og definisjoner, ble det også inkludert engelske oversettelser av uttrykkene. Dette ble gjort blant annet for å vise til det nærmeste uttrykket brukt i andre studier. Man kan kanskje se for seg at uttrykk på to språk kan supplere hverandre og redusere behovet for nærmere definisjon, men vi opplevde det motsatte – to til dels ulike uttrykk på ulike språk skapte usikkerhet og bidro ytterligere til stadig finlesing av definisjonene. Konklusjonen i etterkant var at adferdsuttrykkene kun bør benyttes på ett språk ved scoring, for å unngå forvirring.

Ved første øyekast framstod det unødvendig men uproblematisk å inkludere den engelskspråklige betegnelsen på hvert adferdsuttrykk når vi som norskspråklige skulle utføre QBA scoring. Etter hvert ble vi klar over at slett ikke alle de norske og engelske ordene for samme adferdsuttrykk kan anses å være synonyme eller direkte oversettelser (Cambridge University Press & Assessment, 2024).

“Calm” brukes gjerne om en emosjonell ro, “rolig” kan like gjerne dreie seg om å oppholde seg i ro.

“Hesitant” betyr “nølende”, ikke “avventende” slik det ble definert her (motvillig, tilbaketrukket og årvåken).

“Uneasy” beskriver en emosjonell eller mental uro, mens “urolig” kan tolkes å beskrive enten en emosjonell uro, en stadig motorisk aktivitet eller eventuelt en kombinasjon av begge. Betegnelsen “nøytral” opplevde vi gav lite mening under observasjon av kattene. Vi tenker at “likegyldig” ville være et mer forståelig uttrykk og en presis oversettelse av “indifferent”.

Fra et vitenskapelig synspunkt, vil bruk av to språk fordre at alle brukerne av verktøyet har like ferdigheter i begge språk for at deres scoringer kan inngå i samme analyse.

## **7.5 Fordeler og ulemper med metoden**

### **7.5.1 Fordeler ved bruk av QBA**

**Høy reliabilitet:** Som tidligere nevnt viser studier der QBA har blitt tatt i bruk generelt høy reliabilitet. Både ved bruk av FCP og FL har man generelt sett en høy grad av samsvar mellom observatører (Clarke et al., 2016). En erfaren observatør vil sannsynligvis kunne bruke QBA i en velferdsvurdering. Også uerfarne observatører er vist å kunne bruke metoden reliabelt. (Fleming et al., 2016) hevder at grunnen til at observatører kan oppnå pålitelige resultater uten særlig tidligere erfaring er at selve observasjonen av dyret er i stor grad uavhengig fra subjektive vurderinger. Dette betyr at QBA kan være et pålitelig verktøy for å oppnå enighet i hva som observeres blant observatører av ulik erfaring. Ved bruk av metoden til en reell velferdsvurdering vil man fortsatt alltid bruke en erfaren observatør. Selv om reliabiliteten fortsatt er høy hos de uerfarne, kan det hende deres oppfattelse ikke fanger opp det fulle bildet av adferden like godt som en trent, erfaren observatør (Fleming et al., 2016).



**Helhetlig forståelse gjennom kontekstuell innsikt:** QBA brukes i håp om å få en helhetlig forståelse av velferd gjennom adferd ved å ikke bare vurdere spesifikke handlinger, men også kvalitative aspekter av holdning, bevegelser og uttrykk. Ved å fokusere på kvalitative aspekter tillater QBA observatører å vurdere konteksten der adferden forekommer. Man gjør en vurdering av kattens individualitet, personlighet og emosjoner. Når man heller ser på helhetsbildet av kattens adferd enn enkeltvis uttrykk, gir det rom for å komme med en subjektiv vurdering. Observatører med gode forkunnskaper om katt kan ha godt av dette fordi de kan fange opp nyanser og detaljer i adferd man kanskje ikke ville vurdert ved en rent kvantitativ metode. Dette er viktig for å forstå kattens velferd. Målet er å ikke bare vise fravær av negativ velferd, men også å fange opp positive emosjoner. Dette kan gi et mer komplett bilde av dyrets opplevelse. (Wemelsfelder, 2007)

**Anvendbarhet:** I forhold til en del kvantitative tester brukt for huskatter (Siegford et al., 2003; Travnik & Sant'Anna, 2021; Travnik et al., 2022), tar denne metoden betydelig kortere tid per katt og krever ingen interaksjon, noe som kommer godt til nytte i et shelter der det kan være mange katter på ett sted som også potensielt kan være redde for mennesker. Dersom metoden utføres uten direkte interaksjon, er den dessuten en hensiktsmessig og trygg måte å vurdere aggressive katter på.

## 7.5.2 Ulemper ved bruk av QBA

**Basert på subjektive vurderinger:** Når man scorer adferdsuttrykkene setter man en strek der man intuitivt tenker, på en 125 mm linje, uten å bevisst velge en verdi. Fordi man bygger på subjektive vurderinger, kan resultatene ved bruk av FL-tilnærmingen også være påvirket av individuelle etiske verdier og forståelse av beskrevne adferdsuttrykk (Fleming et al., 2016). Ulike tolkninger kan føre til ulike scoringer. Dette kan gå ut over reliabiliteten av resultatene. (Ved bruk av FCP kan man se for seg en tilsvarende utfordring - at observatørens etiske verdier og bakgrunn påvirker deres opplevelse av dyrets adferd.)

**Krever opplæring:** For å oppnå pålitelige resultater med FL må observatører få en opplæring i hvordan man skal score. Dersom definisjoner / forklaringer benyttes, vil observatørene også ha behov for en grundig gjennomgang av disse. I noen tilfeller kan uttrykkene man scorer virke kompliserte, eller ha en annen definisjon enn hva man kanskje ville tenkt på når man hører uttrykkene alene. Det er også potensielt nødvendig at man jevnlig må kalibreres i bruk av metoden for å sikre at man har konsistensen man trenger i vurderingene sine.

**Tidkrevende kalibrering/ analyse:** Selv om selve observasjonen bare tar noen minutter kan det være tidkrevende å bruke metoden. Å skrive ned scoringer, måle linjene og føre inn målingene til analyse er en tidkrevende prosess. Kalibrering tar også tid, og manglende hyppig kalibrering kan føre til observer-drift: observatøren vil gi ulike scoringer for de samme adferdene etter en periode. Dette kan begrense skalaen av studier eller prosjekter der det kan virke passende å bruke metoden.

## **7.6 Bruksområder for metoden**

### **7.6.1 Shelterprotokoll for katt**

Resultatene fra vår studie indikerer at QBA-metoden har høy reliabilitet på katt blant forhåndsopplærte observatører. I tillegg indikerer tidligere studier at den korrelerer med mer objektive mål på dyrevelferd. (Travnik & Sant'Anna, 2021) Den vil med stor sannsynlighet kunne inkluderes i en protokoll for katt som tilsvarer Welfare Quality og Shelter Quality som brukes på andre arter, for å vurdere velferdskriteriet "positive emotional state". Det finnes så vidt oss bekjent ingen standardisert shelterprotokoll for katt, men vi opplever at behovet for en slik er stort.

### **7.6.2 Identifisering av negativ velferd og terapibehov**

QBA skiller seg ut som metode ved at den fanger opp positive velferdsindikatorer, men den fremstår også som effektiv for å identifisere negativ velferd, hos katt så vel som andre arter. Uttrykk som "fryktsom", "uroelig" og "frustrert" kan fortelle oss at katten føler ubehag i sin situasjon og at vi bør gjøre noe med dette. Dersom årsaken er sykdom eller skade, oppdager vi at vi bør ta med katten til veterinær. Dersom disse kan utelukkes som årsak, og katten gir uttrykk for frykt og angst, kan den ha behov for terapi i form av f.eks. habituering (tilvenning til oppfattet negativ stimuli). En katt som scorer høyt på "frustrert" har sannsynligvis et behov som den ikke får innfridd. Å bearbeide disse negative sinnsstemningene er en viktig del av å ta vare på katter i oppstalling, som QBA altså kan hjelpe til med. Metoden kan også brukes med jevne mellomrom for å vurdere kattens fremgang, både når det gjelder negativ velferd og generell sosiabilitet og adopterbarhet. Det er dessuten godt nytt for omplasseringsorganisasjoner at samsvaret i videoscoring 1 var høyt blant alle grupper etter en rask gjennomgang av begrepene, fordi det indikerer at metoden ikke nødvendigvis krever lang opplæring.

### **7.6.3 Tilsyn med dyrevelferd**

QBA kan potensielt bli et nyttig supplementært verktøy for Mattilsynet og lignende organer som gjør tilsyn med velferden i kattehold. Metoden kan utføres relativt raskt av en trent observatør og gi informasjon som nåværende kvantitative metoder ikke fanger opp. Siden disse som regel er statlige organer, er det svært viktig at metodene de bruker gir konsekvente resultater. Våre funn indikerer at QBA vil ha god reliabilitet dersom det brukes (fortrinnsvis som del av en samlet protokoll) for å vurdere velferd i kattehold.

#### **7.6.4 Adopterbarhet**

I en studie av Gourkow & Fraser undersøkte man hvilke faktorer som påvirket adopterbarhet hos shelterkatter i Canada. Nye eiere ble spurt hva de la vekt på i valget av katt. Alle de 71 eierne oppga "vennlig mot meg" som en viktig årsak. 69% oppga "vennlig mot andre katter", mens 38% oppga "fryktsom". Uttrykk som nettopp "vennlig" og "fryktsom" blir brukt også i QBA, og en rask intuitiv vurdering ved hjelp av en fixed list kan være med på å bevisstgjøre observatøren om hvilke adjektiver som best beskriver den enkelte katten. Et mulig bruksområde for QBA kan dermed være å kartlegge katters adopterbarhet ved bruk av kunnskap om hva potensielle eiere ser etter, uten at man i dette tilfellet behøver noen påfølgende statistisk analyse. Det er viktig å merke seg at adjektivene som ble brukt her, ikke har de samme tilhørende definisjonene som vi har brukt for adferdsuttrykkene i vår studie. Likevel kan de antas å ha en sterk sammenheng med uttrykk som går ut på det samme.

#### **7.6.5 QBA-app: for en mer tilgjengelig metode**

Wemelsfelder og hennes team ved SRUC er ved skrivende stund i prosessen med å utvikle en QBA-app. (SRUC, u.å.) Denne skal gjøre metoden lettere å bruke og forstå. Dette er spesielt viktig fordi analysene som brukes i etterkant av scoring for å vurdere hvert enkelt dyr eller hver enkelt besetning, er avansert og krever stor forståelse av statistikk. Appen skal kunne utføre dette automatisk. Appen automatiserer også millimetermålingen av skalaene, noe som er veldig tidkrevende å gjøre for hånd. Dette vil forhåpentligvis også gjøre metoden mer tilgjengelig og effektiv for organisasjoner som er frivillig drevet og har stor utskiftning av personale.

## 8. Konklusjon

Resultatene av denne studien peker mot en sterk pålitelighet av QBA-metoden til bruk på katt. Metoden viste høy inter- og intra-observatør-reliabilitet mellom kalibrerte og ikke-kalibrerte dyrepleierstudenter med ulik grad av erfaring med katt for observasjoner av katt i 2 minutter.

De fleste uttrykkene hadde høyt samsvar observatørene imellom, etterfulgt av moderat, mens lavt samsvar var svært sjeldent. Dette gjelder både katter på video og i felt. Resultatene tyder også på at de aller fleste adferdsuttrykkene som ble valgt ut var forståelige og beskrev reell adferd som kan observeres hos katt.

Det ble også funnet høyt samsvar blant dyrepleierstudentene som ikke hadde privat erfaring med katt. Dette kan tyde på at erfaring med arten (utover en grunnleggende innføring) i dette tilfellet ikke er en forutsetning for å kunne bruke metoden, så lenge adferden man skal se etter er forklart på forhånd.

QBA-metoden kan potensielt bli et godt verktøy for tilsynsmyndigheter som vurderer dyrevelferd i kattehold, spesielt som del av en større velferdsprotokoll. Den er også potensielt nyttig til bruk på shelter for katt. Etter en enkelt kalibrering kan den både gi nyttig informasjon som et første øyeblikksbilde ved inntak, og ved jevnlig bruk for å vurdere fremgang med de engstelige kattene. I sin nåværende form gir metoden noe mangelfull informasjon om sosiabilitet, men sammen med andre tester kan den høyst sannsynlig også være nyttig for å vurdere adopterbarhet. QBA som verktøy for analyse av velferd fordrer per nå en stor forståelse av statistikk som de fleste sheltere ikke kan forventes å inneha, men lanseringen av en app vil forhåpentligvis gjøre denne prosessen betydelig lettere og metoden mer tilgjengelig.

## 9. Veien videre

Videre studier er nødvendig for å videreutvikle og kvalitetssikre metoden QBA som verktøy for evaluering av velvære og velferd hos shelter katter. For at metoden skal kunne brukes i en mer helhetlig velferdsprotokoll, vil det være behov for utvikling av en slik for katt, for eksempel basert på Shelter Quality Protocol for hund.

Som nevnt i vår drøfting, erfarte vi at definisjonene (de utdypende forklaringene gitt på de ulike adferdsuttrykkene) ikke nødvendigvis var til hjelp under scoring av shelterkattene. Det er foreløpig uvisst hvorvidt å inkludere definisjoner av adferdsuttrykkene er hensiktsmessig for metodens reliabilitet. En systematisk review av QBA-studier som har brukt "Fixed List"-tilnærmingen med og uten definisjoner vil kunne undersøke om dette utgjør en forskjell, og i så fall om forskjellen er signifikant.

Det kunne også vært interessant å gjøre en studie der man undersøker samsvar i scoring mellom 2 grupper av kalibrerte observatører der den ene gruppen aldri hadde hatt tilgang til "definisjonene".

En tredje videoscoring kunne vært interessant for videre å undersøke effekt av kalibrering og observer drift. Hvordan vil sammenlignbar reliabilitet være mellom videoscoring 2 og en potensiell videoscoring 3 i forhold til mellom videoscoring 1 og videoscoring 2? Intra-observatør-reliabiliteten kan tenkes å bli lavere på grunn av effekten av kalibrering, men ved en tredje videoscoring kunne man fått et tydeligere skille mellom observer drift og kalibrering.

Det bør gjennomføres studier som undersøker hvorvidt og i hvor stor grad observatørens tolkning av kattens adferd og velvære påvirkes av om dyrene observeres i et bur. En slik studie kan for eksempel gjøres med videoscoring av simultane opptak fra innsiden og utsiden av buret.

Som tidligere nevnt virker det ikke å finnes beskrivelser av hvor lang observasjonstid som er mest hensiktsmessig når QBA brukes til å vurdere individer i felt. Mangelen på et tidsaspekt å forholde seg til kan gjøre bruk av metoden mindre konsekvent. Derfor bør observasjonstid undersøkes nærmere, med fokus på gjennomførbarhet samtidig som at det ikke går ut over observatørens evne til å fange opp dyrets adferd.

Systematisk evaluering av velvære ved bruk av QBA kan på sikt kanskje bidra til en dypere forståelse av ulike arters behov.

## 10. Takk til bidragsytere

Takk til våre veiledere Karianne Muri, Randi Oppermann Moe og Solveig Marie Stubsjøen, for god veiledning, oppmuntring og grundige og detaljerte kommentarer og innspill gjennom hele prosessen.

Takk til bibliotekarer Linn, Ingeborg og Beate ved NMBU Veterinærhøgskolens studentbibliotek, for god hjelp med kildeføring og EndNote.

Takk til DOOA ved frivilligkoordinator Kristina Lie-Hagen, samt frivillige på Hjelpesenteret, for filming av videoer og tillatelse til å observere kattene på senteret under kalibreringsprosessen.

Takk til FOD ved daglig leder Anna Lie, for tillatelse til å observere kattene på gården under kalibreringsprosessen og under feltscoringen. Takk også til FOD-ansatt Alexander Mendizabal, for filming av videoer samt særdeles god organisering og oppfølging av besøk til gården.

Takk til våre medstudenter Wictoria Karlsen, Eline Lombnæs og Siril Kjeldedal, for å delta som kontrollgruppe ved begge videoscoringene.

## 11. Summary

**Title:** A study of inter- and intra-observer-reliability of qualitative behaviour assessment of cats

**Authors:** Syver Vik Enger, Tobias Humlekjær Palmgren & Sarah Wilberg

**Supervisor:** Karianne Muri, førsteamanuensis II, Institutt for produksjonsdyrmedisin, NMBU veterinærhøgskolen

**Co-supervisors:** Randi Oppermann Moe, professor, Institutt for produksjonsdyrmedisin, NMBU veterinærhøgskolen & Solveig Marie Stubsjøen, seniorforsker, Seksjon husdyr, vilt og velferd, Veterinærinstituttet

In this study, three veterinary nursing students tested the inter- and intra-observer-reliability of the Qualitative Behaviour Assessment (QBA) method when used on cats. The study used a fixed list of 20 predetermined behavioural descriptors, and each cat was scored on a visual analogue scale (VAS) of 125 mm. The observed cats were shelter cats in the care of animal rescue organizations. They were scored twice using video clips, four months apart. In addition to the authors, a control group of 3 other students from the same class also scored the video clips. Reliability was analysed by supervisors using the statistics software Stata. Generally, inter-observer-reliability was moderate-to-high for most descriptors, regardless of observer grouping. Negative descriptors showed lower concordance than positive ones, but not significantly so. Intra-observer-reliability for each observer was also generally high for all descriptors. In addition, 20 shelter cats were scored in the field by the authors. Inter-observer-reliability in the field was somewhat lower than for the video clips, but was still moderate-to-high for most descriptors.

## 12. Referanser

- Alderton, D. (1996). *Cats*. Eyewitness Handbooks. London: Dorling Kindersley Ltd.
- Andersen, I. L. & Halleraker, J. H. (2023). Dyrevelferd. I: *Store norske leksikon*. Tilgjengelig fra: <https://snl.no/dyrevelferd> (lest 28.11.2023).
- Andreasen, S. N., Wemelsfelder, F., Sandoe, P. & Forkman, B. (2013). The correlation of Qualitative Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. *Applied Animal Behaviour Science*, 143 (1): 9-17. doi: 10.1016/j.applanim.2012.11.013.
- APA. (2023). *APA Dictionary of Psychology*. Tilgjengelig fra: <https://dictionary.apa.org/observer-drift> (lest 05.12.2023).
- Appleby, M. C. & Sandoe, P. (2002). Philosophical debate on the nature of well-being: Implications for animal welfare. *Animal Welfare*, 11 (3): 283-294.
- Appleby, M. C., Olsson, I. A. S. & Galindo, F. (2018). Introduction. I: Appleby, M. C., Olsson, I. A. S. & Galindo, F. (red.) *Animal Welfare*, s. XI-XII. Glasgow: CAB International.
- Arena, L., Barnard, S. & Dalla Villa, P. (2014). *Shelter Quality: Welfare Assessment Protocol for Shelter Dogs*. Tilgjengelig fra: [https://www.izs.it/IZS/Engine/RAServeFile.php/f/pdf\\_publicazioni/ProtocolloShelterQuality\\_EN\\_2016-DEF.pdf](https://www.izs.it/IZS/Engine/RAServeFile.php/f/pdf_publicazioni/ProtocolloShelterQuality_EN_2016-DEF.pdf) (lest 12.01.2024).
- Begum, M. R. & Hossain, M. A. (2022). Validity and reliability of visual analogue Scale (vas) for pain measurement. *Journal of Medical Case Reports and Reviews*, 2 (11).
- Braastad, B. O. (2019). *Katten Atferd og velferd 2*. utg. Bergen: Vigmostad Bjørke.
- Bradshaw, J. W. S., Casey, R. A. & Brown, S. L. (2013). *The Behaviour of the Domestic Cat*. 2. utg. Wallingford: CAB International.
- Butterworth, A., A., M. J., Wielebnowski, N. & Olsson, I. A. S. (2018). Practical Strategies to Assess (and Improve) Welfare. I: Appleby, M. C., Olsson, I. A. S. & Galindo, F. (red.) *Animal Welfare*, s. 232-250. Glasgow: CAB International.
- Cambridge University Press & Assessment. (2024). Cambridge Dictionary. I. Tilgjengelig fra: <https://dictionary.cambridge.org/dictionary/english/>.
- Clarke, T., Pluske, J. R. & Fleming, P. A. (2016). Are observer ratings influenced by prescription? A comparison of Free Choice Profiling and Fixed List methods of Qualitative Behavioural Assessment. *Applied Animal Behaviour Science*, 177: 77-83. doi: 10.1016/j.applanim.2016.01.022.
- Deane, K. & Valentine, A. (2022). Assessment and recording methods tool kit. I: Rendle, M., Hinde-Megarity, J. (red.) *BSAVA Manual of Practical Veterinary Welfare*, s. 18-70. Gloucester: British Small Animal Veterinary Association.
- Driscoll, C. A., Clutton-Brock, J., Kitchener, A. C. & O'Brien, S. J. (2009). The Taming of the cat. Genetic and archaeological findings hint that wildcats became housecats earlier--and in a different place--than previously thought. *Sci Am*, 300 (6): 68-75.
- Duncan, I. J. H. (2020). The philosophical and biological evolution of feelings in well-being. I: McMillan, F. D. (red.) *Mental health and well-being in animals*, s. 21-32. Wallingford: CAB International.
- EFSA Panel on Animal Health and Welfare. (2012). Statement on the use of animal-based measures to assess the welfare of animals. *EFSA Journal*, 10 (6): 2767. doi: 10.2903/j.efsa.2012.2767.
- Fleming, P. A., Clarke, T., Wickham, S. L., Stockman, C. A., Barnes, A. L., Collins, T. & Miller, D. W. (2016). The contribution of qualitative behavioural assessment to

- appraisal of livestock welfare. *Animal Production Science*, 56 (10): 1569-1578. doi: 10.1071/An15101.
- Fogle, B. (2006). *Damms store bok om hunder*. 2. utg. Oslo: N. W. Damm & Sønn AS. Foreningen Norske Etologer. (u.å.). *Etologi og dyrevelferd - Dyrevelferd*. Tilgjengelig fra: <https://etologi.no/> (lest 10.09.2023).
- Fraser, D., Weary, D. M., Pajor, E. A. & Milligan, B. N. (1997). A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare*, 6 (3): 187-205.
- Fraser, D. (1999). Animal ethics and animal welfare science: bridging the two cultures. *Applied Animal Behaviour Science*, 65 (3): 171-189. doi: Doi 10.1016/S0168-1591(99)00090-8.
- Fraser, D. (2008). *Understanding Animal Welfare: The Science in its Cultural Context*. UFAW/Blackwell series. Oxford: Wiley-Blackwell.
- Gourkow, N. & Fraser, D. (2006). The effect of housing and handling practices on the welfare, behaviour and selection of domestic cats (*Felis sylvestris catus*) by adopters in an animal shelter. *Animal Welfare*, 15 (4): 371-377.
- Irving-Pease, E. K., Ryan, H., Jamieson, A., Dimopoulos, E. A., Larson, G. & Frantz, L. A. F. (2019). Paleogenomics of Animal Domestication. I: Lindqvist C., R. P. (red.) *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, s. 225-272. Cham: Springer International Publishing.
- Johannessen, C. S., Madsen, H. & Larsen, M. W. (2020). *Kvalitativ atferdsvurdering av hund – samsvar mellom atferdsobservasjoner utført av dyrepleierstudenter*. Ås: Norges miljø- og biovitenskapelige universitet. Tilgjengelig fra: <https://hdl.handle.net/11250/2658030> (lest 20.08.2023).
- Keeling, L. J., Rushen, J. & Duncan, I. J. H. (2018). Understanding Animal Welfare. I: Appleby, M. C., Olsson, I. A. S. & Galindo, F. (red.) *Animal Welfare*, s. 16-35. Glasgow: CAB International.
- Leyhausen, P. (1979). *Cat Behaviour The predatory and Social Behavior of Domestic and Wild Cats*. Garland Series in Ethology. New York: Garland STPM Press.
- McMillan, F. D. & Yeates, J. W. (2020). The problems with well-being terminology. I: McMillan, F. D. (red.) *Mental health and well-being in animals*, s. 8-20. Wallingford: CAB International.
- Mellor, D. J. & Beausoleil, N. J. (2015). Extending the 'Five Domains' model for animal welfare assessment to incorporate positive welfare states. *Animal Welfare*, 24 (3): 241-253. doi: 10.7120/09627286.24.3.241.
- Mellor, D. J. (2017). Operational Details of the Five Domains Model and Its Key Applications to the Assessment and Management of Animal Welfare. *Animals (Basel)*, 7 (8). doi: 10.3390/ani7080060.
- Mellor, D. J., Beausoleil, N. J., Littlewood, K. E., McLean, A. N., McGreevy, P. D., Jones, B. & Wilkins, C. (2020). The 2020 Five Domains Model: Including Human-Animal Interactions in Assessments of Animal Welfare. *Animals (Basel)*, 10 (10). doi: 10.3390/ani10101870.
- Michigan State University. (u.å.). *The Five Freedoms: A history lesson in animal care and welfare*. Tilgjengelig fra: [https://www.canr.msu.edu/news/an\\_animal\\_welfare\\_history\\_lesson\\_on\\_the\\_five\\_freedoms](https://www.canr.msu.edu/news/an_animal_welfare_history_lesson_on_the_five_freedoms) (lest 15.11.2023).
- Moe, R. O. (2016). Dyrevelferd hos fjørfe. I: Bagley, M. F. (red.) *Fjørfeboka*, s. 209-236. Bergen: Vigmostad & Bjørke AS.
- Norges forskningsråd. (2005). *Forskningsbehov innen dyrevelferd i Norge*. Tilgjengelig fra: <https://www.forskningsradet.no/siteassets/publikasjoner/1108644079320.pdf> (lest 01.04.2024).



- PAW. (u.å.). *Qualitative Behaviour Assessment: quantifying the emotionally expressive qualities of animal behaviour*. Tilgjengelig fra: <https://www.positiveanimalwelfare.net/research/qba/> (lest 10.11.2023).
- Phythian, C., Michalopoulou, E., Duncan, J. & Wemelsfelder, F. (2013). Inter-observer reliability of Qualitative Behavioural Assessments of sheep. *Applied Animal Behaviour Science*, 144 (1-2): 73-79. doi: 10.1016/j.applanim.2012.11.011.
- Reips, U. D. & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods*, 40 (3): 699-704. doi: 10.3758/Brm.40.3.699.
- Siegford, J. M., Walshaw, S. O., Brunner, P. & Zanella, A. J. (2003). Validation of a temperament test for domestic cats. *Anthrozoos*, 16 (4): 332-351. doi: Doi 10.2752/089279303786991982.
- Sordo, L., Breheny, C., Halls, V., Cotter, A., Tornqvist-Johnsen, C., Caney, S. M. A. & Gunn-Moore, D. A. (2020). Prevalence of Disease and Age-Related Behavioural Changes in Cats: Past and Present. *Veterinary Sciences*, 7 (3). doi: ARTN 85 10.3390/vetsci7030085.
- SRUC. (u.å.). *Qualitative Behaviour Assessment app*. Tilgjengelig fra: <https://www.sruc.ac.uk/research/research-areas/animal-behaviour-welfare/qualitative-behaviour-assessment-app/> (lest 21.02.2024).
- Stubsjøen, S. M., Moe, R. O., Bruland, K., Lien, T. & Muri, K. (2020). Reliability of observer ratings: Qualitative behaviour assessments of shelter dogs using a fixed list of descriptors. *Veterinary and animal science*, 10: 100145-100145. doi: 10.1016/j.vas.2020.100145.
- Tranøy, K. E. & Ore, Ø. (2020). René Descartes. I: *Store norske leksikon*. Tilgjengelig fra: [https://snl.no/René\\_Descartes](https://snl.no/René_Descartes) (lest 05.01.2024).
- Travnik, I. C. & Sant'Anna, A. C. (2021). Do you see the same cat that I see? Relationships between Qualitative Behaviour Assessment and indicators traditionally used to assess temperament in domestic cats. *Animal Welfare*, 30 (2): 211-223. doi: 10.7120/09627286.30.2.211.
- Travnik, I. C., Machado, D. S. & Sant'Anna, A. C. (2022). Do you see the same cat that I see? Inter- and intra-observer reliability for Qualitative Behaviour Assessment as temperament indicator in domestic cats. *Animal welfare*, 31 (3): 319-327. doi: 10.7120/09627286.31.3.004.
- Tuytens, F. A. M., de Graaf, S., Heerkens, J. L. T., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E. & Ampe, B. (2014). Observer bias in animal behaviour research: can we believe what we score, if we score what we believe? *Animal Behaviour*, 90: 273-280. doi: 10.1016/j.anbehav.2014.02.007.
- Universities of Glasgow & Edinburgh Napier. (2015). *Glasgow Composite Measure Pain Scale: CMPS - Feline*. Tilgjengelig fra: [https://www.aprvt.com/uploads/5/3/0/5/5305564/cmp\\_feline\\_eng.pdf](https://www.aprvt.com/uploads/5/3/0/5/5305564/cmp_feline_eng.pdf) (lest 09.04.2024).
- Welfare Quality Network. (u.å.). *Assessment Protocols*. Tilgjengelig fra: <https://www.welfarequalitynetwork.net/es-es/informes/assessment-protocols> (lest 10.12.2023).
- Wemelsfelder, F. (1997). The scientific validity of subjective concepts in models of animal welfare. *Applied animal behaviour science*, 53 (1): 75-88. doi: 10.1016/S0168-1591(96)01152-5.
- Wemelsfelder, F. (2007). How animals communicate quality of life: the qualitative assessment of behaviour. *Animal Welfare*, 16 (Supplement): 25-31.

Yeates, J. (2022). Animal ethics and welfare. I: Rendle, M. & Hinde-Megarity, J. (red.) *BSAVA Manual of Practical Veterinary Welfare*, s. 1-17. Gloucester: British Small Animal Veterinary Association.

Zaiontz, C. (u.å.). *Kendall's Coefficient of Concordance (W)*. Tilgjengelig fra: <https://real-statistics.com/reliability/interrater-reliability/kendalls-w/> (lest 23.11.2023).

## 13. Vedlegg

### Vedlegg 1: Samtykkeerklæring FOD



Foreningen for omplassering av dyr  
Enebakkveien 866  
1290 Oslo

#### Samtykkeerklæring

Prosjektet «Kvalitativ atferdsvurdering: en ny metode for å vurdere katters velferd» er et samarbeid mellom forskergruppene i dyrevelferd ved Veterinærinstituttet og NMBU Veterinærhøgskolen. Hovedmålet er å kvalitetssikre kvalitativ atferdsvurdering som en metode for å vurdere velferd hos katt. Prosjektet ledes av Solveig Marie Stubsjøen ([solveig-marie.stubsjoen@vetinst.no](mailto:solveig-marie.stubsjoen@vetinst.no)), fagansvarlig dyrevelferd ved Veterinærinstituttet.

Med denne erklæringen gir undertegnede samtykke til at innsamlet data/materiale (dvs. videoopptak av katter som uttrykker ulik atferd og direkte observasjoner av katters atferd) kan brukes til studentenes bacheloroppgave, undervisning, publisering i fagartikler og videre forskning ved Veterinærinstituttet og NMBU Veterinærhøgskolen. All data som samles inn vil bli anonymisert.

Innsamlet data vil være tilgjengelig så lenge det er faglig relevant og vil kun brukes til formålene beskrevet ovenfor.

Navn: AR Lie

Sted, dato: Oslo 05.12.23

Veterinærinstituttet  
Arboretveien 57, 1433 Ås  
Postboks 64, 1431 Ås  
Besøk: Elizabeth Stephansens vei 1, 1433 Ås

TLF 23 21 60 00  
EPOST [postmottak@vetinst.no](mailto:postmottak@vetinst.no)  
WEB [vetinst.no](http://vetinst.no)

ORG. NR 970 955 623 MVA  
BANKKORTID 7694 05 12030

## Vedlegg 2: Samtykkeerklæring DOOA



Veterinærinstituttet  
Norwegian Veterinary Institute

Dyrebeskyttelsen Norge Oslo og Akershus  
Mosseveien 226  
1169 Oslo

### Samtykkeerklæring

Prosjektet «Kvalitativ atferdsvurdering: en ny metode for å vurdere katters velferd» er et samarbeid mellom forskergruppene i dyrevelferd ved Veterinærinstituttet og NMBU Veterinærhøgskolen. Hovedmålet er å kvalitetssikre kvalitativ atferdsvurdering som en metode for å vurdere velferd hos katt. Prosjektet ledes av Solveig Marie Stubsjøen ([solveig-marie.stubsjoen@vetinst.no](mailto:solveig-marie.stubsjoen@vetinst.no)), fagansvarlig dyrevelferd ved Veterinærinstituttet.

Med denne erklæringen gir undertegnede samtykke til at innsamlet data/materiale (dvs. videoopptak av katter som uttrykker ulik atferd og direkte observasjoner av katters atferd) kan brukes til studentenes bacheloroppgave, undervisning, publisering i fagartikler og videre forskning ved Veterinærinstituttet og NMBU Veterinærhøgskolen. All data som samles inn vil bli anonymisert.

Innsamlet data vil være tilgjengelig så lenge det er faglig relevant og vil kun brukes til formålene beskrevet ovenfor.

Navn:

Sted, dato: \_\_\_\_\_ Oslo, 21.11.23 \_\_\_\_\_

Veterinærinstituttet  
Arboretveien 57, 1433 Ås  
Postboks 64, 1431 Ås  
Besøk: Elizabeth Stephansens vei 1, 1433 Ås

TLF 23 21 60 00  
EPOST [postmottak@vetinst.no](mailto:postmottak@vetinst.no)  
WEB [vetinst.no](http://vetinst.no)

ORG. NR 970 955 623 MVA  
BANKKORTID 7694 05 12030

**Vedlegg 3: VAS scoringsskjema (nedskalert)**

NAVN: \_\_\_\_\_ KATT NR.: \_\_\_\_\_ DATO: \_\_\_\_\_

Rolig Min. \_\_\_\_\_ Max.

Ukomfortabel Min. \_\_\_\_\_ Max.

Selvsikker Min. \_\_\_\_\_ Max.

Fryktsom Min. \_\_\_\_\_ Max.

Vennlig Min. \_\_\_\_\_ Max.

Frustrert Min. \_\_\_\_\_ Max.

Avslappet Min. \_\_\_\_\_ Max.

Anspent Min. \_\_\_\_\_ Max.

Sosial Min. \_\_\_\_\_ Max.

Stresset Min. \_\_\_\_\_ Max.

Interessert Min. \_\_\_\_\_ Max.

Avventende Min. \_\_\_\_\_ Max.

Tilfreds Min. \_\_\_\_\_ Max.

---

Opphisset	Min.	_____	Max.
Tillitsfull	Min.	_____	Max.
Urolig	Min.	_____	Max.
Fredfull	Min.	_____	Max.
Nøytral	Min.	_____	Max.
Aktiv	Min.	_____	Max.
Undersøkende	Min.	_____	Max.