

**Universität für Bodenkultur Wien**

Department für Nachhaltige Agrarsysteme  
Institut für Nutztierwissenschaften



# **Einfluss der Tageszeit auf die Ergebnisse des Qualitative Behaviour Assessment bei Milchkühen**

## **Masterarbeit**

vorgelegt von  
Bernhard Schwed, Bakk.Biol.

Betreuung durch

Univ. Prof. Dr. Christoph Winckler  
Dipl. Biol. Anke Gutmann  
DI Lukas Tremetsberger

Wien, März 2014

# INHALTSVERZEICHNIS

<b>1 Einleitung .....</b>	<b>4</b>
1.1 Ziel der Arbeit, Forschungsfragen.....	4
<b>2 Literaturübersicht .....</b>	<b>6</b>
2.1 Erfassung des Wohlergehens von Tieren.....	6
2.1.1 Quantitative und qualitative Methoden.....	7
2.1.2 Free Choice Profiling (FCP).....	8
2.1.3 Welfare Quality®.....	9
2.1.4 Fixed Term Assessment (FT).....	11
2.2 Tagesrhythmik der Rinder.....	15
<b>3 Material und Methoden .....</b>	<b>16</b>
3.1 Untersuchungsbetriebe.....	16
3.2 Untersuchungsaufbau.....	19
3.3 Voruntersuchung, um die Videoclplänge und -aufbereitung als mögliche Einflussfaktoren der QBA-Bewertung zu testen.....	19
3.3.1 Material und Methoden.....	20
3.3.2 Ergebnisse.....	22
3.3.3 Diskussion.....	23
3.4 Hauptuntersuchung zum Einfluss der Tageszeit.....	24
3.5 Datenaufbereitung und statistische Auswertung.....	24
<b>4 Ergebnisse .....</b>	<b>27</b>
4.1 Übereinstimmung zwischen BeobachterInnen.....	27
4.2 Effekt der Beobachtungszeit auf die Ergebnisse der qualitativen Verhaltensbeurteilung.....	29
4.2.1 Vollständiger Datensatz.....	29
4.2.2 Reduzierter Datensatz.....	34
<b>5 Diskussion .....</b>	<b>40</b>
5.1 Methodik.....	40
5.1.1 Versuchsdesign.....	40
5.1.2 Vorteile, Nachteile und Grenzen der Videobewertung.....	42
5.2 Ergebnisse.....	44

5.2.1 BeobachterInnenübereinstimmung beim Qualitative Behaviour Assessment.....	44
5.2.2 Einfluss der Tageszeit auf die qualitative Verhaltensbeurteilung.....	46
<b>6 Schlussfolgerung und Ausblick .....</b>	<b>51</b>
<b>7 Zusammenfassung .....</b>	<b>52</b>
<b>8 Danksagung .....</b>	<b>54</b>
<b>9 Literaturverzeichnis .....</b>	<b>55</b>
<b>10 Anhang .....</b>	<b>60</b>
<b>11 Ehrenerklärung .....</b>	<b>62</b>

*„Die Tiere empfinden wie der Mensch Freude und Schmerz, Glück und Unglück;  
sie werden durch dieselben Gemütsbewegungen betroffen wie wir.“*

Charles Darwin

# 1 Einleitung

Das Wohlergehen von Nutztieren gewinnt in Europa für Konsumenten zunehmend an Bedeutung (Miele und Parisi, 2001). Durch dieses gesellschaftliche Anliegen wurde es notwendig, verlässliche, auf wissenschaftlichen Erkenntnissen basierende Systeme zu entwickeln, welche die Beurteilung des Wohlergehens von landwirtschaftlichen Nutztieren auf Praxisbetrieben ermöglichen (Blokhuys et al., 2008).

Die im Rahmen des EU-Projektes Welfare Quality<sup>®</sup> erarbeiteten Erhebungsprotokolle stellen ein solches System zur Beurteilung von Tierwohlergehen dar. Neben einer Reihe von quantitativen Beurteilungen am Tier, einer Ressourcen-Checkliste und einem Managementfragebogen ist die Methode des Qualitative Behaviour Assessment (Qualitative Verhaltensbeurteilung, kurz QBA) enthalten (Wemelsfelder et al., 2009c).

Grundsätzlich erfolgt die Beurteilung des Wohlergehens von Milchkühen auf Praxisbetrieben nach einem im Bewertungsprotokoll festgelegten Zeitplan. Dabei wird die Bewertung mittels Qualitative Behaviour Assessment am frühen Morgen durchgeführt (Wemelsfelder et al., 2009c).

Bei Rindern unterliegt das Verhalten einer Tagesrhythmik. Die höchste Aktivität besteht in der Morgen- und Abenddämmerung (Winckler, 2009, 89), wohingegen der späte Vormittag und der Nachmittag zum Ruhen und Wiederkauen genutzt werden (Houpt, 2011, 80). Auf dem bei Rindern vorliegenden, hier beschriebenen Tagesrhythmus gründet sich die Annahme, dass das am Tier beobachtete Verhalten zu unterschiedlichen Tageszeiten variiert, was bisher für QBA nicht untersucht wurde.

## **1.1 Ziel der Arbeit, Forschungsfragen**

Das Ziel der vorliegenden Arbeit war es, für die Methode des gemäß WQ Protokoll angewandten QBA zu prüfen, ob es unter Praxisbedingungen im Tagesverlauf zu unterschiedlichen Ergebnissen in der Bewertung von Milchkuhverhalten mittels QBA kommt.

Darüber hinaus wurde überprüft, ob es zu unterschiedlichen Effekten kommt, wenn die Ergebnisse zum einen gemäß des WQ Protokolls, zum anderen mit einer unabhängigen Hauptkomponentenanalyse berechnet werden.

Folgende Forschungsfragen wurden im Rahmen der Masterarbeit überprüft:

- a) Variieren die Bewertungsergebnisse des QBA in Abhängigkeit von der Tageszeit der Beobachtung signifikant voneinander?
- b) Welche Effekte treten auf, wenn die Beobachtung zu unterschiedlichen Tageszeiten durchgeführt wird?
- c) Kommt es zu unterschiedlichen Effekten, wenn die Ergebnisse gemäß WQ Protokoll bzw. mit einer unabhängigen Hauptkomponentenanalyse berechnet werden?

## 2 Literaturübersicht

### 2.1 Erfassung des Wohlergehens von Tieren

Die Forschung im Bereich Animal Welfare ist eine junge, aber mittlerweile gut etablierte wissenschaftliche Disziplin. Die Zahl der Publikationen steigt rasant an (Blokhuis et al., 2008). Der englische Begriff des „Animal Welfare“ wird im Deutschen häufig mit „Wohlergehen“ übersetzt (Sachser, 2009; Kirchner, 2012).

Broom formulierte 1986 eine der am häufigsten zitierten Definitionen, indem er Wohlergehen als die Fähigkeit des Tieres mit seiner Umwelt zurechtzukommen beschrieb (Broom, 1986).

In die wissenschaftliche Beurteilung von Wohlergehen fließen drei Konzepte mit ein:

- (1) ungestörte biologische Funktion
- (2) emotionaler Zustand
- (3) Artgemäßes Verhalten (Fraser, 2003).

Es herrscht unter Wissenschaftlern Konsens darüber, dass eine breite Vielfalt an Methoden angewandt werden muss, um das Wohlergehen von Tieren zu beurteilen (Canali und Keeling, 2009).

Grundsätzlich gibt es zwei Herangehensweisen um das Wohlergehen von Nutztieren wissenschaftlich zu erfassen: ressourcenbezogene und tierbezogene Parameter (Main et al., 2003).

Bei einem ausschließlich ressourcenbezogenen Erhebungssystem können dem Beobachter wichtige Informationen verborgen bleiben. Whay et al. (2003) gaben zu bedenken, dass allein die Bereitstellung guten Managements und einer guten Haltungsumwelt noch nicht garantieren, dass ein Tier gesund und fit ist und einen hohen Standard an Wohlergehen erreicht. In der wissenschaftlichen Forschung wird deshalb zumeist eine Kombination aus ressourcenbezogenen und tierbezogenen Messungen eingesetzt (Main et al., 2003).

Bei der tierbezogenen Erhebung kann wiederum zwischen quantitativen und qualitativen Methoden unterschieden werden.

### 2.1.1 Quantitative und qualitative Methoden

Zu den quantitativen Methoden, das Wohlergehen von Rindern zu messen, zählen unter anderem die Ausweichdistanz, Lahmheitsbeurteilung sowie Abliegeverhalten (Wemelsfelder et al., 2009c). Ein Problem besteht darin, dass durch verschiedene quantitative Methoden bei Betriebsbesuchen separat ein Teilbereich des Wohlergehens betrachtet wird, jedoch diese Informationen anschließend zu einem Gesamtbild des Wohlergehens zusammengefügt werden sollen (Wemelsfelder et al., 8/2001a). Wemelsfelder (1997) argumentiert, dass die Untersuchung von subjektiven Zuständen von Tieren wie Wohlergehen und Leiden geradezu nach der Anwendung einer qualitativen Methode verlangen, weil es sich dabei um dynamische Zustände handelt. Sie postuliert, dass Wohlergehen und Leiden über das Verhalten der Tiere ausgedrückt werden. Doch Verhalten kann nicht durch die Überprüfung von einzelnen Bewegungen oder die Verfassung zu einem bestimmten Augenblick erfasst werden. Vielmehr ist es nötig, den Verlauf des Verhaltens über eine gewisse Zeitspanne hinweg zu beobachten, bevor dessen dynamische Qualitäten ersichtlich werden. Eine Methode, die diese Voraussetzungen erfüllt, ist das QBA (Wemelsfelder und Lawrence, 2001b).

Diese Methode wurde im vergangenen Jahrzehnt am Scottish Agricultural College entwickelt (Wemelsfelder et al., 2009c). Mit ihrem subjekt-basierten Ansatz hat Françoise Wemelsfelder, die maßgeblich die Erforschung dieser Methode vorangetrieben hat, im Bereich der wissenschaftlichen Erhebung von tierischem Wohlergehen Neuland betreten (Spolder et al., 2003).

QBA basiert auf der Integration des Ausdrucks bzw. der Körpersprache der Tiere (Rousing und Wemelsfelder 2006; Wemelsfelder et al., 2001). Zur Beschreibung des am Tier beobachteten Verhaltens dienen Begriffe wie „ruhig“, „aggressiv“, „sozial“ oder „gleichgültig“. Dabei werden auch subtile Details des Verhaltens miteinbezogen, die mit den herkömmlichen quantitativen Methoden zum Teil gar nicht erfasst werden können, mitunter aber entscheidend in der Beurteilung des Wohlergehens sein können (Wemelsfelder et al., 2001a).

Beim Qualitative Behaviour Assessment ist die Frage nach dem „Was tut das Tier?“ von untergeordneter Relevanz. Vielmehr konzentriert sich der Beobachter auf das „Wie tut das Tier etwas?“ (z.B. wie frisst es), was durch die Beobachtung



und Interpretation der Körpersprache beurteilt werden kann (Wemelsfelder, 1997; Wemelsfelder et al., 2001a). Wemelsfelder (1997) betont, dass bei diesem Ansatz das Tier als Subjekt betrachtet werden muss: Das Tier, als Ursprung jeglichen Verhaltens verstanden, darf nicht in einzelne mechanische Teile zergliedert werden, sondern das Verhalten muss aus der Perspektive des Individuums beurteilt werden.

Dass beim Qualitative Behaviour Assessment dem Beobachter eine aktivere, interpretierende Rolle zufällt, als dies bei den traditionell quantitativen Methoden der Fall ist, war lange Zeit ein Hindernis für die wissenschaftliche Akzeptanz dieser Methode (Wemelsfelder, 1997). Häufig standen Wissenschaftler der Methode misstrauisch gegenüber, da sie fürchteten, Anthropomorphismus würde zu falschen Ergebnissen führen (Wemelsfelder et al., 2001a). Eine Reihe von Studien hat gezeigt, dass Ergebnisse des Qualitative Behaviour Assessment gut mit denen traditioneller Methoden der Verhaltensforschung korrelieren (Minero et al., 2009; Rousing und Wemelsfelder, 2006).

Es wurden zwei verschiedene Methoden des Qualitative Behaviour Assessment entwickelt: Beim „Free Choice Profiling“ finden die Beobachter ihre eigenen Begriffe um das Verhalten zu beurteilen, wohingegen bei den sogenannten „Fixed Terms“ schon vorgegebene Begriffe verwendet werden.

### **2.1.2 Free Choice Profiling (FCP)**

Als man vor gut fünfzehn Jahren erste Untersuchungen zur Erfassung des Wohlergehens von Tieren mittels QBA durchführte, wurde zunächst ausschließlich die Methode des Free Choice Profiling (FCP) verwendet (Wemelsfelder et al., 2001a).

FCP wurde aus der Lebensmittelwissenschaft übernommen und die Eignung dieser Methode zur Bewertung von Tierverhalten wurde in mehreren Studien überprüft (Wemelsfelder et al., 2009d; Bokkers et al., 2012; Rutherford et al., 2012; Andreasen et al., 2013). FCP besteht aus zwei Phasen: In der ersten Phase findet jeder Beobachter individuell Begriffe, die das Verhalten der beobachteten Tiere bestmöglich widerspiegeln. In Phase zwei gewichtet der Beobachter seine in

Phase 1 gewählten persönlichen Begriffe durch eine Markierung auf einer jedem Begriff zugeordneten Skala (Wemelsfelder et al., 2001a).

Beim FCP ist ein Minimum von zehn Beobachtern Voraussetzung, damit eine statistische Auswertung durchführbar ist. Dies macht die Methode des FCP zu einer komplizierten und zeitintensiven Prozedur. Somit kann FCP nicht für die On-farm-Bewertung verwendet werden (Wemelsfelder et al., 2009c).

### **2.1.3 Welfare Quality®**

2004 startete das europäische Forschungsprojekt Welfare Quality®. Welfare Quality® basierte auf einem Konsortium von 44 Forschungsinstitutionen in 17 Ländern, deren Aufgabe es war, in gemeinsamer Arbeit einen europäischen Forschungsraum für den Bereich Tierschutz hinsichtlich der Qualität der Produkte zu etablieren (Knierim und Winckler, 2009). Welfare Quality® trug dem öffentlichen Anliegen nach besserem Wohlergehen von Nutztieren Rechnung und wollte durch transparente Qualität dem Konsumenten die Identifikation von Produkten aus artgerechter Tierhaltung ermöglichen (Wemelsfelder et al., 2009c). Ein zentrales Ziel dieses Projekts war es, ein zuverlässiges Erhebungssystem mit Schwerpunkt auf tierbezogene Messgrößen zu entwickeln, das sich auf wissenschaftlich solide Daten stützt und praktisch durchführbar ist (Blokhuis et al., 2003). Dazu wurden On-farm-Bewertungsprotokolle für folgende Nutztierarten entwickelt: Geflügel (Welfare Quality®, 2009a), Schweine (Welfare Quality®, 2009b) und Rinder (Welfare Quality®, 2009c). Die Beurteilung gemäß Welfare Quality®-Protokoll basiert auf einer Kombination der drei oben genannten Konzepte ungestörte biologische Funktion, emotionaler Zustand und artgemäßes Verhalten.

Welfare Quality®-Erhebungsprotokolle sind nach „Welfare-Prinzipien“ sowie „Welfare-Kriterien“ gegliedert. Dabei beinhaltet jedes der vier „Welfare-Prinzipien“ eine Schlüsselfrage des Wohlergehens. Die „Welfare-Prinzipien“ sind: gute Fütterung, gute Haltung, gute Gesundheit und artgemäßes Verhalten. Sie entsprechen den Fragen:

- (1) Sind die Tiere ordentlich mit Futter und Wasser versorgt?
- (2) Sind die Tiere ordentlich untergebracht?

(3) Sind die Tiere in einem guten Gesundheitszustand?

(4) Können die Tiere ihre natürlichen Verhaltensmuster ausleben?

Jedes der vier „Welfare-Prinzipien“ wiederum beinhaltet zwei bis vier voneinander unabhängige „Welfare-Kriterien“. Diese sind in Tabelle 1 dargestellt. Somit werden alle Bereiche tierischen Wohlergehens abgedeckt. Um das Wohlergehen von Nutztieren auf Praxisbetrieben gemäß Welfare Quality®-Erhebungsprotokoll zu evaluieren, ist für jedes der zwölf „Welfare-Kriterien“ zumindest eine Messgröße vorgesehen. Das „Welfare-Kriterium“ „Positive emotionale Befindlichkeit“ wird mittels QBA erhoben. Grundsätzlich sind die Messgrößen so gewählt, dass nicht das Fachwissen eines Veterinärmediziners bzw. eines Nutztierethologen nötig ist, um die Erhebungen durchzuführen.

**Tabelle 1: Die „Welfare Quality® Prinzipien“ und „Welfare Quality® Kriterien“ als Basis des Welfare Quality®-Erhebungsprotokolls (Welfare Quality®, 2009c)**

<b>„Welfare Quality® Prinzipien“</b>	<b>„Welfare Quality® Kriterien“</b>
<i>Gute Fütterung</i>	1 Abwesenheit von anhaltendem Hunger
	2 Abwesenheit von anhaltendem Durst
<i>Gute Haltung</i>	3 Liegekomfort
	4 Thermischer Komfort
	5 Bewegungsfreiheit
<i>Gute Gesundheit</i>	6 Abwesenheit von Verletzungen
	7 Abwesenheit von Krankheit
	8 Abwesenheit von Schmerzen aufgrund des Managements
<i>Artgemäßes Verhalten</i>	9 Äußerung des Sozialverhaltens
	10 Äußerung von anderem Verhalten
	11 Gute Mensch-Tierbeziehung

## 12 Positive emotionale Befindlichkeit

Nachdem alle Erhebungen durchgeführt sind, wird aus den Messergebnissen für jedes „Welfare-Kriterium“ eine Punktezahl von 0 bis 100 ermittelt. Dabei meint 0 die schlechteste Situation, in der sich ein Tier befinden kann, bei der keine weitere Verschlechterung des Wohlergehens mehr möglich ist. Eine Punktezahl von 100 entspricht der besten Situation, die vorgefunden werden kann, für die keine weitere Verbesserung des Wohlergehens vorstellbar ist. Die Scores für die „Welfare-Kriterien“ wiederum werden mittels Choquet-Integral zusammengefügt, um die Punktezahl für die vier „Welfare-Prinzipien“ zu erhalten. Die Ergebnisse aller „Welfare-Prinzipien“ schließlich werden dazu herangezogen, den Betrieb in eine von vier Kategorien von Wohlergehen einzuteilen (Welfare Quality®, 2009c). Diese Kategorien sind:

Exzellent:	Das Wohlergehen der Tiere ist auf dem höchsten Niveau.
Gehoben:	Das Wohlergehen der Tiere ist gut.
Akzeptabel:	Das Wohlergehen der Tiere entspricht den Minimumanforderungen bzw. liegt darüber.
Nicht klassifiziert:	Das Wohlergehen der Tiere ist niedrig und wird deshalb als nicht ausreichend betrachtet (Welfare Quality®, 2009c).

### 2.1.4 Fixed Term Assessment (FT)

Da die im Rahmen des Welfare Quality®- Projekts erarbeiteten Erhebungsprotokolle darauf ausgelegt sind, dass sie auf Praxisbetrieben einfach anzuwenden sind, war es nötig, für das QBA ein Bewertungssystem mit vorgegebenen Begriffen zu entwickeln. Dies ermöglicht eine standardisierte Auswertung der Bewertung mittels QBA (Knierim und Winckler, 2009).

Um eine Rating-Skala für die Beurteilung des Verhaltens von Milchkühen zu entwickeln, wurde in Italien eine Studie, die 22 Milchviehgruppen umfasste, durchgeführt. Insgesamt vier geschulte BeobachterInnen führten das Qualitative Behaviour Assessment direkt auf den Betrieben durch und hatten zur Bewertung

29 vorgegebene Begriffe zur Verfügung. Die Entscheidung für diese Begriffe fiel auf der Basis vorausgegangener Studien mit Rindern sowie durch Konsultation von Rinderexperten. Ziel war es, die vorgegebenen Begriffe unter Praxisbedingungen hinsichtlich der interindividuellen Beobachterübereinstimmung zu testen (Wemelsfelder et al., 2009c).

Die ursprünglich 29 Begriffe wurden im Rahmen der Auswertung auf 20 Begriffe reduziert, um die Bewertung einfacher zu gestalten. Aus der Liste entfernt wurden Begriffe, die (1) alle Beobachter nicht für aussagekräftig hielten, (2) deren Bedeutung keine klare Relevanz für tierisches Wohlergehen haben, oder (3) deren Bedeutung Begriffen auf der Liste ähnlich war (Wemelsfelder et al., 2009c).

Gleich in mehreren Arbeiten betonen Wemelsfelder et al. (2001a; 2001b; 2009d), dass sie die Methode des FCP gegenüber den Fixed Terms (FT) bevorzugen. FCP, so Wemelsfelder et al. (2001a), gibt den BeobachterInnen die Möglichkeit alle wahrgenommenen Details des Verhaltens selbstständig in Begriffe zusammenzufassen, die nach deren Meinung am besten den Ausdruck des Verhaltens der Tiere beschreiben. Bei Verwendung der FT hingegen sind die BeobachterInnen in ihrer Beurteilung weit weniger frei und durch die vorgegebenen Begriffe in ihrer Entscheidung beeinflusst, was insgesamt zu einer weniger feinsinnigen Beurteilung führt (Wemelsfelder et al., 2009d).

Um nun die Frage zu klären, ob anstelle des zeitaufwändigen FCP auch fixe vorgegebene Begriffe zur On-Farm-Beurteilung von Milchviehverhalten geeignet sind, wurde 2011 auf der Universität für Bodenkultur, Wien eine Masterarbeit zu diesem Thema durchgeführt (Müllner, 2011). In dieser Studie wurde eine Videobeurteilung mit 12 BeobachterInnen durchgeführt. Es wurden jeweils 20 Videoclips, die Einzeltiere bzw. Gruppen von Kühen zeigten, sowohl mit der Methode des FCP als auf mittels FT beurteilt. Müllner konnte zeigen, dass beide Methoden zu sehr ähnlichen Ergebnissen führen. Somit zeigte die Studie, dass fixe Begriffe für On-Farm-Beurteilungen von Milchkühen eingesetzt werden können, wenn FCP aufgrund des hohen Zeit- und Personalaufwandes nicht möglich ist (Müllner, 2011).

Das Welfare Quality<sup>®</sup>-Erhebungsprotokoll für Milchkühe verfügt über 20 vorgegebene Begriffe, die jeweils auf einer Skala beurteilt werden. Neben jedem

der 20 Begriffe befindet sich eine 125 Millimeter lange durchgehende Linie, die am linken Ende durch „Minimum“ und am rechten durch „Maximum“ definiert ist. Dabei meint „Minimum“, dass an diesem Punkt keines der beobachteten Tiere ein Verhalten mit dieser Qualität zeigt, wohingegen bei „Maximum“ die ausgedrückte Qualität des Verhaltens, die durch den jeweiligen Begriff angezeigt wird, bei sämtlichen Tieren ausgeprägt vorhanden ist. Die BeobachterInnen beurteilen das Tierverhalten, indem sie bei allen 20 Begriffen mit einem Strich den für sie richtigen Punkt auf der Linie markieren (Welfare Quality®, 2009c).

Der „Kriterien-Score“ für das „Welfare-Kriterium“ „Positive emotionale Befindlichkeit“ ist die Punktezahl, die ein Betrieb aufgrund der Werte für die unterschiedlichen Begriffe erreicht und die in das Prinzip „Artgemäßes Verhalten“ bei der Berechnung des Welfare Quality®-Gesamtscore einfließt.

Die Auswertung der Bewertungsbögen erfolgt, indem der Abstand von der Minimum-Markierung auf der linken Seite der Skala zu dem vom Beobachter gesetzten Strich in Millimetern gemessen wird. Die Begriffe sind im Welfare Quality®-Protokoll unterschiedlich gewichtet. Begriffe, die schlechtes Wohlergehen anzeigen sind negativ gewichtet. Begriffe, die positives Wohlergehen anzeigen sind positiv gewichtet. Zur Berechnung des Punktwertes aus der QBA-Bewertung wird eine im Welfare Quality®-Protokoll gegebene Berechnungsformel verwendet: Dazu werden die gemessenen Distanzen (von 0 bis 125 mm) der 20 Begriffe, die jeder Betrieb erreicht, in einen Index umgerechnet:

Index (I):  $-3,40496 + \text{Summe der gewichteten Begriffe}$

Schließlich wird aus dem Index der Score für die QBA-Bewertung (WQ QBA Score) berechnet:

Score (bei  $I \leq 0$ ) =  $-(10 \times I) - (1,25 \times I^2)$

Score (bei  $I \geq 0$ ) =  $50 + (11,667 \times I) - (0,55556 \times I^2)$

Der „Kriterien-Score“ für die Bewertung mittels QBA sieht Punktwerte von 0 bis 100 vor, auch wenn in der Berechnung prinzipiell höhere oder niedrigere Werte auftreten können (Welfare Quality®, 2009c).

Der im Welfare Quality®-Protokoll berechnete „Kriterien-Score“ verwendet vorgegebene Gewichtungen, die auf Referenzdatensätzen beruhen. Es besteht aber auch die Möglichkeit, die Ergebnisse aus der QBA-Bewertung mittels FT unabhängig von vorgegebenen Gewichtungen zu analysieren:

Mit Hilfe von Multivariaten Analysemethoden werden Datensätze, die, wie der Name sagt, mehrere Variablen enthalten, untersucht. Im Falle des QBA mittels FT ergibt sich durch die Anzahl der Begriffe ein Datensatz von 20 Variablen. Die Hauptkomponentenanalyse oder englisch Principal Component Analyse (PCA) ist ein derartiges multivariates Analyseverfahren. Sie hat zum Ziel, die in einer großen Anzahl von Ursprungsvariablen enthaltene Information zu vereinfachen, indem sie zu wenigen neuen Dimensionen zusammengefasst wird und dabei möglichst wenig Information verliert (McGarigal et al., 2000, 20ff). Die PCA liefert genau so viele Dimensionen wie Variablen in die Berechnung einfließen. Es werden in weiterer Folge jedoch nur so viele Dimensionen verwendet, wie nötig, um einen angemessen großen Teil der Gesamtvariation zu erklären (Fowler et al., 1998, 214). Jede dieser neuen Dimensionen ist durch eine lineare Kombination der Ursprungsvariablen definiert und wird auch als Hauptkomponente (englisch Principal Component) bezeichnet (McGarigal et al., 2000, 23). Mit der Komponentenladung wird die Korrelation zwischen jeder einzelnen Variable und den Hauptkomponenten angegeben. Die Komponentenladungen weisen somit darauf hin, wie eng eine Variable mit einer Dimension verbunden ist. Somit erfolgt auch die Interpretation der Dimensionen auf Basis der Komponentenladungen: Ist das Ausmaß der Komponentenladung hoch (wenn es sich an  $\pm 1$  annähert), trägt die Dimension beinahe dieselbe Information wie die Variable. Im Gegenzug dazu haben die beiden wenig gemeinsam, wenn die Komponentenladung nahe bei Null liegt. Häufig werden Werte größer oder kleiner 0,50 als signifikant bezeichnet, wobei es hier keine absoluten Grenzen gibt (McGarigal et al., 2000, 50ff).

Ein wesentlicher Bestandteil der Komponentenanalyse ist die Interpretation der Dimensionen. Sie wird nicht von einem Computerprogramm durchgeführt, sondern erfolgt als intuitiver Prozess durch den Forscher selbst (McGarigal et al., 2000, 53f). Bei der Interpretation der Dimensionen werden Variablen mit geringer Komponentenladung ignoriert, Variablen mit hoher Komponentenladung hervorgehoben (Fowler et al., 1998, 217f). Anschließend versucht der Forscher im

Muster der Komponentenladungen eine Bedeutung zu finden, die er mit einem Namen benennt. Dieser Begriff soll in größtmöglicher Weise das reflektieren, was die einzelnen Komponentenladungen auf der Dimension repräsentieren (McGarigal et al., 2000, 53).

## **2.2 Tagesrhythmik der Rinder**

Grundsätzlich ist es für jede Verhaltensstudie wichtig, einen Tageszeitpunkt festzulegen, an dem die Beobachtung stattfinden soll. Tiere sind ganz offenkundig nicht den ganzen Tag über gleich aktiv. Daher hängt die wahrgenommene Aktivität des Tieres sehr stark vom Zeitpunkt ab, zu dem man es beobachtet. Martin und Bateson (2007, 51ff) weisen überdies darauf hin, dass abgesehen vom Aktivitätsniveau auch die Art des Verhaltens variieren kann. Harcourt (1978) zum Beispiel beobachtete in einer Feldstudie an Gorillas, dass die Art der sozialen Interaktionen unter den Tieren in Abhängigkeit von deren Aktivitätsniveau variierte. Auch bei Rindern unterliegt das Verhalten einem spezifischen Tagesrhythmus. Die Futteraufnahme geschieht vorwiegend tagsüber. Da Rinder dämmerungsaktiv sind, erfolgen die umfangreichsten Fressperioden im Morgengrauen und zur Abenddämmerung. Zwar wurde die meiste Forschung zum Futteraufnahmeverhalten von Wiederkäuern an Weidetieren durchgeführt, doch wurde gezeigt, dass auch bei Rindern in Stallhaltung ein ähnlicher Rhythmus beobachtet werden kann (DeVries et al., 2003). Bei Stallhaltung können Milchkühe durch Melken, frische Futtergaben oder Nachschieben von weggeschobenem Futter zum Fressen motiviert werden, wobei frisches Futter die Futteraufnahme am stärksten stimuliert und Nachschieben von weggeschobenem Futter die geringste Wirkung zeigt (DeVries und von Keyserlingk, 2005; DeVries et al., 2003).

Außerdem steigt zu den Fresszeiten die Häufigkeit agonistischer Interaktionen. (Winckler, 2009, 85ff). Darunter sind aggressive Verhaltensweisen sowie die Reaktionen auf aggressives Verhalten zu verstehen, bei dem es sich meistens um Ausweichen handelt.

Nach Houpt (2011, 80) werden der späte Vormittag und der Nachmittag vorrangig zum Ruhen und Wiederkauen genutzt.



## **3 Material und Methoden**

### ***3.1 Untersuchungsbetriebe***

Aufgrund der Kooperation mit einem Dissertationsprojekt hielten alle in die Untersuchung einbezogenen Betriebe mindestens 25 Milchkühe in Liegeboxenlaufställen. Da die Wirtschaftsweise auf die Forschungsfrage keinen Einfluss hat, konnten grundsätzlich sowohl konventionelle als auch biologisch wirtschaftende Betriebe an der Studie teilnehmen. Tatsächlich aber wurden alle zehn in die Untersuchung aufgenommenen Betriebe konventionell bewirtschaftet. Sechs Untersuchungsbetriebe lagen im Bezirk Amstetten in Niederösterreich, ein Betrieb im Bezirk Scheibbs in Niederösterreich, ein weiterer im Bezirk Waidhofen an der Ybbs in Niederösterreich und zwei Betriebe im Bezirk Steyr-Land im Bundesland Oberösterreich. Alle zehn Betriebe führten eine Ganzjahresstallfütterung mit Gras- bzw. Gras- und Maissilage sowie Heu und Krafffutter durch. Drei Untersuchungsbetriebe verabreichten das Futter als Totalmischration, ein Betrieb fütterte eine aufgewertete Mischration. Fünf Betriebe verfügten über zumindest eine Krafffutterstation. Bei vier Betrieben war zumindest ein Teil der Lauffläche als Betonspaltenboden ausgeführt, nur ein Betrieb verfügte über einen betonierten Auslauf. Diese und weitere Betriebsdaten wurden in Tabelle 2 zusammengefasst.

**Tabelle 2: Betriebsdaten**

<b>Betriebscode</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Herdenalter in Jahren	4,7	4,1	5,2	4,2	5,0
Stalldurchschnitt in kg Milch	8243	8888	4900	7736	8446
Kuhzahl Betrieb	38	33	34	29	32
Kuhzahl bei Laktierenden	38	28	35	25	31
Anzahl Liegeboxen	42	49	40	23	31
Anzahl FP	39	29	40	25	33
KF-Transponder	0	0	0	1	1
Transponder geschützt*	0	0	0	1	1
Anzahl Tränke	2	4	3	3	4
Lauffläche: Fressgang	Beton	Flächenelemente	Beton	Gummilaufflächen	planbefestigt
Fressgangbreite	3,4m	3,5m	4,50m	4,2m	3,30m
Lauffläche: Laufgang	Beton	Beton	Beton	Gummilaufflächen	planbefestigt
Laufgangbreite	2,50m	?	2m	3m	?
Anzahl Sackgassen	0	3	1	0	0
Bürste	1 mech.	1 rotierende + 1 mech.	1 rotierende	1 rotierende + 1 mech.	1 mech.
Klima	Warmstall	Zwischenstufe **	Zwischenstufe	Warmstall	Warmstall
Laufhof	nein	ja	nein	ja	nein

\* 1...Transponder komplett geschützt, 0...Seite und Kopfbereich geschützt. \*\* Haltungsform zwischen Warmstall und Kaltstall, mit permanenten Öffnungen in den Stallwänden

<b>Betriebscode</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
Herdenalter in Jahren	5,0	4,4	4,7	4,0	5,2
Stalldurchschnitt in kg Milch	7696	8172	6188	9711	8303
Kuhzahl Betrieb	32	35	27	37	34
Kuhzahl bei Laktierenden	28	30	25	30	31
Anzahl Liegeboxen	38	30	25	30	30
Anzahl FP	33	30	24	30	25
KF-Transponder	2	0	1	0	1
Transponder geschützt*	0	0	1	0	0
Anzahl Tränke	2	2	5	2	2
Lauffläche: Fressgang	Gummilauffl.	Beton	Gummi. + Fl.elemente	Flächenelemente	Gussasphalt
Fressgangbreite	4m	3,50m	3,27m	2,70-3,10m	3,40m
Lauffläche: Laufgang	Gummilaufflächen	Beton	Flächenelemente	Flächenelemente	Gussasphalt
Laufgangbreite	3m	?	2,6m	?	2,50m
Anzahl Sackgassen	0	0	1	2	0
Bürste	1 rotierende	1 rotierende	1 rotierende	1 rotierende + 1 mech.	1 mech. + 1 rotierende
Klima	Warmstall	Zwischenstufe**	Warmstall	Zwischenstufe	Zwischenstufe
Laufhof	nein	nein	nein	nein	nein

\* 1...Transponder komplett geschützt, 0...Seite und Kopfbereich geschützt. \*\* Haltungform zwischen Warmstall und Kaltstall, mit permanenten Öffnungen in den Stallwänden

### **3.2 Untersuchungsaufbau**

Es wurden im Zeitraum vom 9. November 2011 bis zum Frühjahr 2012 Videoaufnahmen auf insgesamt 10 Praxisbetrieben angefertigt. Um zu prüfen, ob die Tageszeit einen Einfluss auf das QBA-Ergebnis hat, wurde das Tierverhalten auf jedem Untersuchungsbetrieb zu drei unterschiedlichen Zeitpunkten (nach dem Morgenmelken um ca. 8 Uhr, vormittags um ca. 11 Uhr, am frühen Nachmittag um ca. 13 Uhr) auf Video aufgezeichnet. Für die Videoaufzeichnung wurde ein Panasonic HDC-SD99 HD Camcorder sowie eine Sony DCR-SX30 Handycam verwendet. Das gesammelte Videomaterial umfasste pro Betrieb drei Aufnahmen von jeweils 20 Minuten. Da der gesamte Stallbereich nicht von einem einzigen Standort einzusehen war, wurde der Stall vor Untersuchungsbeginn in maximal sechs Segmente unterteilt (4 bis 6 Segmente) und innerhalb des Segments ein Standpunkt für die Beobachtung festgelegt, der außerhalb des für die Tiere zugänglichen Bereiches lag und der bei späteren Aufzeichnungen bestehen blieb. Um das gesamte Segment gut überblicken zu können, wurde die Kamera gegebenenfalls mittels ausfahrbarem Stativ in einer erhöhten Position montiert. Die Aufnahmedauer je Segment richtete sich an der Gesamtaufnahmezeit von 20 min und der Anzahl an Segmenten. An jedem Segment war die Aufnahmedauer gleich lange und die Segmentaufnahmen folgten unmittelbar aufeinander. Während der gesamten Beobachtung verhielten sich alle im Stall anwesenden Personen möglichst ruhig und versuchten, die Tiere nicht zu beeinflussen.

### **3.3 Voruntersuchung, um die Videocliplänge und -aufbereitung als mögliche Einflussfaktoren der QBA-Bewertung zu testen**

Schon häufig wurden in vorausgegangenen Studien zum QBA Videoclips zur Bewertung herangezogen. Der klare Vorteil liegt hier im Gegensatz zur Live-Bewertung in der Möglichkeit, die Videoclips öfters und in unterschiedlicher Weise zu analysieren. Für bestimmte Forschungsfragen ist es nötig, dass eine Gruppe von BeobachterInnen dieselbe Situation bewertet. In diesem Fall ist es häufig leichter zu bewerkstelligen, die BeobachterInnen für eine oder mehrere

Bewertungseinheiten in einem Vorlesungssaal zu versammeln, als jedes Mal gemeinsam Betriebe zu besuchen.

Die Dauer der gezeigten Videoclips variierte in den vergangenen Studien zum QBA: Es wurden Videos von 4 Minuten Länge (Wemelsfelder und Lawrence, 2001b; Wemelsfelder et al., 2001a), 3-minütige Videoclips (Wemelsfelder et al., 2009d) bzw. in Studien zum Sozialverhalten 1-minütige Clips (Müllner, 2011; Rousing und Wemelsfelder, 2006) bewertet.

Für die vorliegende Arbeit stellte sich somit die Frage, welche Länge der Videoclips angemessen ist. Um den BeobachterInnen einen möglichst guten Gesamteindruck von der Situation im Stall zu geben, war es erforderlich, zumindest ein zweites Stallsegment im Video zu zeigen. Sollten die zwei Segmente jedoch in einem Videoclip mit Segmentwechsel oder in zwei Videoclips gezeigt werden? Somit ergab sich im Zuge der Arbeit eine weitere Forschungsfrage, die in einem Vortest beantwortet werden sollte:

Kommt es bei unterschiedlicher Videocliplänge und -aufbereitung zu signifikanten Abweichungen in der Bewertung mittels QBA?

### **3.3.1 Material und Methoden**

Von sechs Untersuchungsbetrieben wurden jeweils zwei repräsentative Kamerapositionen gewählt. Die Wahl der Segmente erfolgte nach folgenden drei Prinzipien:

- 1) Am Beginn des Videoclips müssen mindestens fünf Tiere gut sichtbar sein;
- 2) Alle relevanten Stallbereiche wie Futtertisch, Laufgänge, Kraffutterautomaten und Liegeboxen müssen gezeigt werden;
- 3) Es dürfen keine Überschneidungen von zwei aufeinander folgenden Segmenten gezeigt werden.

Der Beginnzeitpunkt der Videoausschnitte wurde standardisiert. Vom ausgewählten Segment wurden die ersten 30 Sekunden verworfen und die folgenden zwei Minuten verwendet. In der Vorbewertung wurde eine Variante von zwei aneinander gehängten 2-minütigen Videoclips von insgesamt vier Minuten

Länge gegen eine Variante der getrennt beurteilten kurzen 2-minütigen Videoclips getestet.

An der Vorbewertung nahmen sieben BeobachterInnen, zwei Männer und fünf Frauen, teil. Zwei BeobachterInnen waren MasterstudentInnen, vier waren NutztierwissenschaftlerInnen. Außerdem nahm eine technische Mitarbeiterin teil. Alle BeobachterInnen waren mit dem grundlegenden Verhaltensrepertoire von Rindern vertraut. Fünf BeobachterInnen hatten bereits praktische Erfahrung in der QBA-Bewertung, zwei BeobachterInnen hatten ausschließlich theoretische Kenntnisse, hatten jedoch QBA selbst noch nicht angewendet.

Der Bewertung der Videoclips ging ein etwa einstündiges QBA-Training voraus. Dieses bestand aus einer Einführung in die Methode des QBA sowie einer praktischen Anleitung zur Durchführung von QBA mittels Fixed Terms. Im Rahmen dieses Trainings wurden die Bedeutung aller 20 Begriffe besprochen und Abweichungen im Verständnis der Begriffsbedeutung diskutiert. Anschließend wurde ein 4-minütiges Probevideo mit Segmentwechsel zu Trainingszwecken bewertet und die Bewertungsergebnisse wurden untereinander verglichen und diskutiert.

In der Voruntersuchung wurden zuerst sechs 4-minütige Videoclips mit Segmentwechsel und anschließend zwölf kurze 2-minütige Clips mittels Fixed Terms bewertet, wobei in beiden Videoclipvarianten exakt dasselbe Bildmaterial zu sehen war. Alle Videos wurden in Farbe und mit Ton gezeigt. Dazu wurde die Welfare Quality®-Bewertungsskala für die qualitative Verhaltensbeurteilung von Milchrindern herangezogen (Welfare Quality®, 2009c). Die Begriffe wurden zuvor ins Deutsche übertragen.

Die Auswertung der Bewertungsbögen aus der Videobeobachtung erfolgte durch Abmessen des Abstandes von der Minimum-Markierung auf der linken Seite der Skala zu dem von der beobachtenden Person gesetzten Strich auf der Linie mittels Lineal. Die Abstände wurden in Millimetern gemessen, wobei in Fällen, bei denen der Strich zwischen zwei Millimetermarkierungen auf dem Lineal lag, der untere Wert gewählt wurde. In einem Excel-Datenblatt wurden aus den Daten sämtlicher Bewertungsbögen für alle an der Videobeobachtung teilnehmenden Personen Datenmatrizen erstellt.

Gemäß der im Welfare Quality®-Protokoll gegebenen Berechnungsformel wurde für jede Einzelbewertung der „Kriterien Score“ für das „Welfare-Kriterium“ „Positive emotionale Befindlichkeit“ berechnet. Dieser Wert wird in der vorliegenden Arbeit als Welfare Quality® QBA-Score bzw. WQ QBA-Score abgekürzt.

Die Berechnung der WQ QBA-Scores beruht auf Gewichtungen bzw. Koeffizienten für jeden Begriff, die auf einem Referenz-Datensatz beruhen und in einer PCA berechnet wurde. Die Gewichtungen der Begriffe, die in Abhängigkeit des zu Grunde liegenden Datensatzes variieren, beeinflussen möglicherweise die WQ QBA-Scores unterschiedlich. Deshalb wurde mit den Rohdaten eine PCA durchgeführt, aus der eigene Scores für die ersten beiden Dimensionen resultierten (PCA QBA-Scores). Diese Berechnungen wurden im Statistikprogramm PASW Statistics 18 durchgeführt.

Komponentenladungen spiegeln die Korrelation zwischen den Begriffen und den Hauptkomponenten wider. Begriffe mit hohen Ladungen (Werte > 0,5) wurden ermittelt, denn sie zeigen an, dass der Begriff eng mit der Dimension verbunden ist (McGarigal et al., 2000, 53f). Das Komponentendiagramm wurde interpretiert und die Achsen wurden mit einem Überbegriff bezeichnet.

Mittels gemischtem Modell für wiederholte Messungen wurde der Einfluss der Videocliplänge auf die QBA-Bewertung überprüft. Als abhängige Variablen aus der QBA-Bewertung wurden Dimension 1 der Hauptkomponentenanalyse sowie der WQ QBA-Score herangezogen. Dabei gingen die zwei unterschiedlichen Videocliplängen als Innersubjektfaktoren und die einzelnen Videoclips als Zwischensubjektfaktoren in das Modell ein. Um die beiden Videoclipvarianten miteinander zu vergleichen, wurden die beiden Ergebnisse der 2-minütigen Videoclips, die jeweils den selben Betrieb zeigten, gemittelt.

### **3.3.2 Ergebnisse**

Dimension 1 der Hauptkomponentenanalyse erklärte 37% der Varianz, Dimension 2 erklärte 14% der Varianz, wobei Dimension 1 mit dem Überbegriff „Stimmung“ und Dimension 2 mit dem Überbegriff „allgemeine Aktivität“ bezeichnet wurden. Bezieht man zusätzlich noch Dimension 3 ein, wurden 64% der Varianz erklärt. In

der Auswertung der gesammelten Daten wurde kein signifikanter Effekt der Videocliplänge gefunden (PCA QBA-Score Dimension 1:  $P = 0,912$ , WQ QBA-Score:  $P = 0,414$ ).

### **3.3.3 Diskussion**

Die vorliegende Arbeit orientiert sich am Welfare Quality®-Erhebungsprotokoll. Die Beurteilung von 4-minütigen Videoclips mit Einstellungswechsel kommt der On-Farm-Beurteilung näher als die Kurzvideos von zwei Minuten Länge. Für die Beantwortung der Forschungsfrage nach dem Einfluss der Tageszeit auf das QBA sollte die beobachtende Person einen möglichst guten Eindruck der Situation erhalten. Möglicherweise sind zwei Minuten dafür nicht ausreichend.

Außerdem ist die Wahl der 4-minütigen Videoclips für die Beobachterinnen und Beobachter als weniger ermüdend zu betrachten, da nur die Hälfte an Bewertungsbögen ausgefüllt werden müssen. Dauern Bewertungssitzungen zu lange, kann die Fähigkeit der beobachtenden Person zur akkuraten Beurteilung aufgrund von Ermüdung und Konzentrationsverlust eingeschränkt sein (Martin und Bateson, 2007, 122).

Ein möglicher Nachteil bei der Beurteilung von Videos mit Einstellungswechsels ist es, dass besonders nach längerem Beobachten von unterschiedlichen Videoclips die Beobachterinnen und Beobachter den zweiten Teil stärker im Gedächtnis behalten als den ersten und daher der erste Teil des Videoclips weniger in die Bewertung einfließt als der zweite Teil. Dazu kann jedoch bemerkt werden, dass dasselbe Phänomen auch bei einer Live-Beobachtung auftreten kann, vor allem dann, wenn der Stall in bis zu sechs Segmente eingeteilt wird.

Da es in der QBA-Bewertung von Videoclips unterschiedlicher Länge und Aufbereitung zu keinen signifikanten Abweichungen kam und eindeutig mehr Argumente für die 4-minütigen Videoclips mit Segmentwechsel sprachen, wurden diese anstatt der kurzen 2-minütigen Clips in der Hauptbewertung herangezogen.



### **3.4 Hauptuntersuchung zum Einfluss der Tageszeit**

Für die Hauptuntersuchung standen für jeden der drei unterschiedlichen Zeitabschnitte (morgens, später Vormittag, früher Nachmittag) zehn Videoclips á 4 min zur Verfügung, die den Beobachterinnen und Beobachtern in Farbe und mit Ton gezeigt wurden.

Die BeobachterInnengruppe umfasste 13 Personen, zehn Frauen und drei Männer. Bei den beteiligten Personen handelte es sich um MasterstudentInnen, NutztierwissenschaftlerInnen und eine technische Mitarbeiterin. Wie in der Vorbewertung war es eine Voraussetzung, als BeobachterIn mit dem grundlegenden Verhaltensrepertoire von Milchkühen vertraut zu sein.

Sechs BeobachterInnen hatten bereits praktische Erfahrung in der QBA-Bewertung von Rindern, die übrigen sieben BeobachterInnen hatten noch keine praktische Erfahrung in der Bewertung von Rinderverhalten mittels QBA.

Daher fand vor der Bewertung der Videoclips ein etwa einstündiges Training statt, das inhaltlich wie beim Vortest aufbereitet war. Am Ende des Trainings stand wiederum die Beurteilung eines 4-minütigen Probevideos mittels Fixed Terms und eine anschließende Diskussion der Beurteilungsergebnisse.

### **3.5 Datenaufbereitung und statistische Auswertung**

#### *Datenaufbereitung*

Insgesamt füllten die Beobachterinnen und Beobachter 390 Bewertungsbögen aus. Die Auswertung der Bewertungsbögen erfolgte analog zur Voruntersuchung. In einem Excel-Datenblatt wurden die Daten sämtlicher Bewertungsbögen gesammelt und der WQ QBA-Score berechnet.

Zusätzlich zu den WQ QBA-Scores wurden auf Basis der Rohdaten mittels PCA im Statistikprogramm PASW Statistics 18 PCA QBA-Scores berechnet, um beurteilen zu können, ob es bei Anwendung von zwei unterschiedlichen Berechnungsmethoden zu qualitativen Abweichungen in den Ergebnissen kommt.

Die Komponentenmatrix wurde graphisch dargestellt. Anschließend erfolgte die Interpretation und Benennung der Dimensionen.

### *Statistische Auswertung*

Mittels gemischtem Modell für wiederholte Messungen wurde der Einfluss der Tageszeit auf die WQ QBA-Scores und PCA QBA-Scores überprüft. Dabei gingen die drei unterschiedlichen Tageszeiten als Innersubjektfaktoren und die zehn Betriebe als Zwischensubjektfaktoren in das Modell ein.

Anschließend wurde eine einfaktorielle ANOVA auf Betriebsebene durchgeführt, um zu überprüfen, ob diese zu unterschiedlichen Tageszeiten verschieden bewertet wurden.

Eine ausreichende Beobachterübereinstimmung ist erforderlich, um auszuschließen, dass Unterschiede zwischen Gruppen auf der unterschiedlichen Bewertung der BeurteilerInnen beruhen (Martin und Bateson 2007, 51ff). In keiner Studie, an der zwei oder mehr BeobachterInnen beteiligt sind, sollte daher die Messung des Grads der Übereinstimmung zwischen zwei oder mehr BeobachterInnen fehlen.

Eine einfache und häufig verwendete Methode, um die Übereinstimmung zwischen BeobachterInnen zu messen, ist die Verwendung von Korrelationen. Beim Spearman's Rangkorrelationskoeffizienten werden die Daten in Ränge konvertiert, bevor der Korrelationskoeffizient berechnet wird. Eng damit verbunden ist der mit dem Rangkorrelationskoeffizienten verwandte Konkordanzkoeffizient  $W$  nach Kendall, der verwendet wird, wenn die Übereinstimmung zwischen mehreren BeobachterInnen ermittelt wird. Letzterer wurde auch in der vorliegenden Arbeit berechnet. Kendall's  $W$  ist eine nicht-parametrische statistische Methode, die den Grad des Zusammenhangs zwischen  $k$  Mengen an Rangreihen angibt (Martin und Bateson 2007, p. 51ff).

Grundsätzlich gibt es keine Regel, welcher Grad an Übereinstimmung nötig ist, doch es gibt eine Richtlinie von Martin und Bateson (2007, 51ff), der wir folgten. Sie unterscheiden fünf Kategorien: geringe Korrelation: Koeffizienten von 0,2 bis 0,4; mäßige Korrelation: Koeffizienten von 0,4 bis 0,7; hohe Korrelation:

Koeffizienten von 0,7 bis 0,9 und sehr hohe Korrelation: Koeffizienten von 0,9 bis 1,0. Auch Meagher (2009) behauptet, dass Korrelationen von 0,7 und darüber für ein akzeptables Ergebnis erforderlich sind.

## 4 Ergebnisse

### 4.1 Übereinstimmung zwischen BeobachterInnen

In der BeobachterInnengruppe, die 13 Personen umfasste, reichten die Kendall's Konkordanzkoeffizienten  $W$  für die PCA QBA-Scores für Dimension 1 in Abhängigkeit von der Tageszeit von 0,29 (gering) bis 0,50 (mäßig).  $W$ -Koeffizienten für die PCA QBA-Scores für Dimension 2 lagen zwischen 0,50 und 0,58 (mäßig). Die Koeffizienten für die WQ QBA-Scores reichten von 0,23 (gering) bis 0,46 (mäßig) (Tabelle 3).

**Tabelle 3: Kendall's Konkordanzkoeffizient  $W$  für 13 BeobachterInnen für die Parameter PCA QBA-Scores für Dimension 1 und Dimension 2 sowie WQ QBA-Score zu verschiedenen Tageszeiten.**

QBA Parameter	Kendall's $W$ Morgen:	Kendall's $W$ Vormittag:	Kendall's $W$ Nachmittag:
PCA QBA-Score Dimension 1	0.37	0.29	0.50
PCA QBA-Score Dimension 2	0.54	0.58	0.50
WQ QBA Score	0.37	0.23	0.46

Die vorliegende Arbeit wurde gemäß den Anforderungen des Welfare Quality<sup>®</sup>-Erhebungsprotokolls angelegt. Eine dieser Anforderungen ist eine ausreichend hohe Übereinstimmung zwischen den BeobachterInnen. Diese wurde im *vollständigen Datensatz* der 13 BeobachterInnen nicht erreicht ( $W$ -Koeffizienten  $< 0,7$ ). Um diejenigen BeobachterInnen zu identifizieren, die in ihrer Bewertung der 30 Videoclips am besten übereinstimmten, wurden Spearman Rangkorrelationen

für alle 13 BeobachterInnen berechnet. Nur diejenigen fünf BeobachterInnen, für die eine Mindestübereinstimmung vorlag ( $r_s > 0,5$ ), blieben im reduzierten Datensatz erhalten. Mit diesem reduzierten Datensatz wurden die Analysen wiederholt.

In Tabelle 4 sind die Ergebnisse des Kendall's  $W$ -Test für den PCA QBA-Score von Dimension 1 und Dimension 2 sowie für den WQ QBA-Score für den reduzierten Datensatz der 5 BeobachterInnen dargestellt. Der Kendall's  $W$ -Koeffizient für den PCA QBA-Score von Dimension 1 reichte in Abhängigkeit von der Tageszeit von 0,58 (mäßig) bis 0,76 (hoch).  $W$ -Koeffizienten für den PCA QBA-Score von Dimension 2 lagen zwischen 0,56 (mäßig) und 0,76 (hoch) und für den WQ QBA-Score lagen die Werte zwischen 0,59 (mäßig) und 0,77 (hoch).

**Tabelle 4: Kendall's Konkordanzkoeffizient  $W$  der QBA Parameter PCA QBA-Scores von Dimension 1 und Dimension 2 sowie WQ QBA-Score zu verschiedenen Tageszeiten aus der Beurteilung von 5 BeobachterInnen.**

<b>QBA Parameter</b>	<b>Morgen: Kendall's <math>W</math></b>	<b>Vormittag: Kendall's <math>W</math></b>	<b>Nachmittag: Kendall's <math>W</math></b>
PCA QBA-Score Dimension 1	0.76	0.58	0.61
PCA QBA-Score Dimension 2	0.75	0.76	0.56
WQ QBA-Score	0.77	0.59	0.66

## 4.2 Effekt der Beobachtungszeit auf die Ergebnisse der qualitativen Verhaltensbeurteilung

### 4.2.1 Vollständiger Datensatz

Im folgenden Abschnitt werden die Ergebnisse der WQ QBA-Scores und PCA QBA-Scores für Dimension 1 und 2, die für die unterschiedlichen Tageszeiten (Morgen, Vormittag und Nachmittag) berechnet wurden, dargestellt. Alle in diesem Abschnitt dargestellten Ergebnisse beziehen sich auf den *vollständigen Datensatz* der BeobachterInnengruppe, die 13 Personen umfasste.

#### WQ QBA-Score

Der durchschnittliche WQ QBA-Score reichte von 16,6 ( $\pm 11,0$  SD) bis 50,5 ( $\pm 18,4$  SD).

Ein signifikanter Einfluss auf den WQ QBA-Score bestand für den Faktor „Tageszeit“ ( $F_{2,240} = 25,278$ ,  $P < 0.001$ ) sowie die Wechselwirkung zwischen „Tageszeit“ und „Betrieb“ ( $F_{18,240} = 3.269$ ,  $P < 0.001$ ).

Bei insgesamt drei Betrieben, „2“, „6“ und „8“, bestand ein signifikanter Unterschied der Bewertung in Abhängigkeit von der Tageszeit ( $P < 0,05$ ). In allen drei Betrieben erreichte der mittlere WQ QBA-Score am Morgen signifikant niedrigere Werte als am Vormittag, wohingegen die Beurteilung des Videos vom Nachmittag nicht signifikant vom Morgen unterschiedlich war (Tabelle 5).

**Tabelle 5: Mittelwert ( $\pm$  SD) WQ QBA-Score für 10 Betriebe beurteilt von 13 BeobachterInnen**

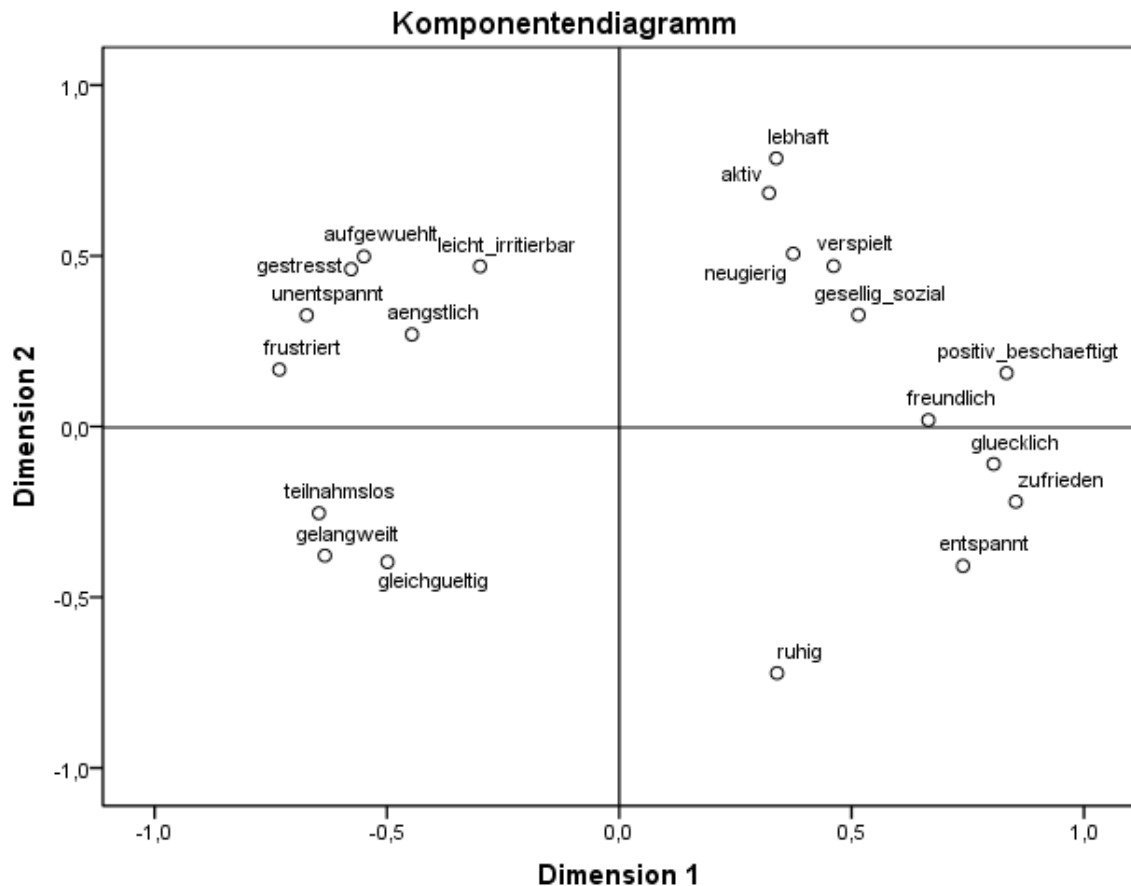
Betrieb	Morgen: Mittelwert ( $\pm$ SD)	Vormittag: Mittelwert ( $\pm$ SD)	Nachmittag: Mittelwert ( $\pm$ SD)
1	45,7 ( $\pm 20,2$ )	38,1 ( $\pm 20,5$ )	38,3 ( $\pm 19,9$ )
2	<b>17,5 (<math>\pm 18,8</math>) a</b>	<b>42,4 (<math>\pm 14,5</math>) b</b>	<b>25,7 (<math>\pm 16,4</math>) a</b>
3	34,5 ( $\pm 21,1$ )	40,4 ( $\pm 17,0$ )	42,5 ( $\pm 17,6$ )
4	19,4 ( $\pm 12,9$ )	25,6 ( $\pm 12,3$ )	21,1 ( $\pm 11,4$ )

5	31,7 (± 17,1)	33,5 (± 16,3)	17,9 (± 20,7)
<b>6</b>	<b>16,6 (± 11,0) a</b>	<b>43,4 (± 15,9) b</b>	<b>28,9 (± 15,0) a</b>
7	30,5 (± 17,6)	37,0 (± 15,6)	30,2 (± 18,9)
<b>8</b>	<b>19,4 (± 11,7) a</b>	<b>35,5 (± 19,1) b</b>	<b>17,7 (± 15,1) a</b>
9	27,4 (± 18,2)	39,9 (± 14,6)	34,5 (± 21,9)
10	33,9 (± 20,5)	50,5 (± 18,4)	49,2 (± 19,9)

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).

### *Hauptkomponentenanalyse*

Der Barlett-Test auf Sphärizität ( $P < 0,001$ ) ergab, dass die Stichprobe für die Durchführung einer Hauptkomponentenanalyse geeignet war. Insgesamt drei Dimensionen erklärten jeweils mehr als 10% der Varianz. Dimension 1 erklärte 35% der Varianz, Dimension 2 erklärte 18% der Varianz. Bezieht man zusätzlich noch Dimension 3 ein, wurden 65% der Varianz erklärt. Davon wurden die beiden ersten Dimensionen interpretiert. Abbildung 1 enthält das Komponentendiagramm der PCA für die Dimensionen 1 und 2. Wie im Komponentendiagramm ersichtlich ist, laden auf Dimension 1 besonders stark die Begriffe 'unentspannt', 'frustriert' und 'teilnahmslos' im negativen Bereich der Achse, sowie 'glücklich', 'zufrieden', 'positiv beschäftigt' im positiven Bereich. Dimension 1 kann mit dem Überbegriff „Stimmung“ interpretiert werden. Die Begriffe 'ruhig', 'gleichgültig', 'entspannt' laden stark im negativen Bereich von Dimension 2, sowie 'lebhaft', 'aktiv', 'neugierig' stark im positiven Bereich dieser Achse. Somit kann Dimension 2 mit dem Begriff „allgemeine Aktivität“ beschrieben werden.



**Abbildung 1: Verteilung der Begriffe im Komponentendiagramm.**

**Komponentenladung gibt die Korrelation der Begriffe mit den Dimensionen 1 und 2 an. Betriebe beurteilt durch 13 BeobachterInnen.**

#### *PCA QBA-Score für Dimension 1*

Für den PCA QBA-Score für Dimension 1 lag für den Faktor „Tageszeit“ ( $F_{2,240} = 17,447$ ,  $P < 0,001$ ) und die Interaktion zwischen „Tageszeit“ und „Betrieb“ ( $F_{18,240} = 3,607$ ,  $P < 0,001$ ) ein signifikanter Einfluss vor.

In der einfaktoriellen ANOVA konnte ein signifikanter Unterschied der PCA QBA-Scores für Dimension 1 in Abhängigkeit von der Tageszeit bei insgesamt drei Betrieben gefunden werden ( $P < 0,05$ ). In zwei Betrieben, „2“ und „6“, war der mittlere PCA QBA-Score von Dimension 1 am Morgen signifikant niedriger als am Vormittag. Bei Betrieb 2 war die Bewertung am Nachmittag nicht signifikant von der Morgenbewertung unterschiedlich und beim Betrieb 6 war die Nachmittagsbewertung, sowohl von der Morgen- als auch von der Nachmittagsbeurteilung, nicht signifikant unterschiedlich. Bei Betrieb 5 war der



Morgen nicht signifikant unterschiedlich vom Vormittag, jedoch gab es einen signifikanten Unterschied zwischen Morgen- und Nachmittagsbeurteilung (Tabelle 6).

**Tabelle 6: Mittelwert ( $\pm$ SD) PCA QBA-Score von Dimension 1 für 10 Betrieben beurteilt von 13 BeobachterInnen**

<b>Betrieb</b>	<b>Morgen: Mittelwert (<math>\pm</math> SD)</b>	<b>Vormittag: Mittelwert (<math>\pm</math> SD)</b>	<b>Nachmittag: Mittelwert (<math>\pm</math> SD)</b>
1	0,79 ( $\pm$ 0,95)	0,23 ( $\pm$ 1,04)	0,22 ( $\pm$ 0,97)
2	<b>-0,75 (<math>\pm</math> 1,12) a</b>	<b>0,33 (<math>\pm</math> 0,74) b</b>	<b>-0,40 (<math>\pm</math> 0,87) a</b>
3	0,29 ( $\pm$ 0,95)	0,61 ( $\pm$ 0,68)	0,43 ( $\pm$ 0,93)
4	-0,79 ( $\pm$ 0,70)	-0,53 ( $\pm$ 0,76)	-0,72 ( $\pm$ 0,77)
5	<b>0,10 (<math>\pm</math> 0,78) a</b>	<b>0,16 (<math>\pm</math> 0,79) a</b>	<b>-0,90 (<math>\pm</math> 1,20) b</b>
6	<b>-0,76 (<math>\pm</math> 0,71) a</b>	<b>0,48 (<math>\pm</math> 0,72) b</b>	<b>-0,14 (<math>\pm</math> 0,67) ab</b>
7	-0,07 ( $\pm$ 0,89)	0,22 ( $\pm$ 0,79)	-0,17 ( $\pm$ 0,87)
8	-0,44 ( $\pm$ 0,89)	0,11 ( $\pm$ 0,87)	-0,73 ( $\pm$ 0,91)
9	-0,17 ( $\pm$ 1,05)	0,28 ( $\pm$ 0,66)	0,33 ( $\pm$ 1,11)
10	0,19 ( $\pm$ 1,14)	0,88 ( $\pm$ 0,86)	0,94 ( $\pm$ 0,93)

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).

#### *PCA QBA-Score für Dimension 2*

Die Varianzanalyse mit Messwiederholung fand für den PCA QBA-Score von Dimension 2 einen signifikanten Effekt der „Tageszeit“ ( $F_{2,240} = 82,229$ ,  $P < 0,001$ ) und der Interaktion zwischen „Tageszeit“ und „Betrieb“ ( $F_{18,240} = 9,963$ ,  $P < 0,001$ ). In der einfaktoriellen ANOVA wurde für die PCA QBA-Scores von Dimension 2 bei insgesamt acht Betrieben ein signifikanter Unterschied der Bewertung in Abhängigkeit von der Tageszeit gefunden ( $P < 0,05$ ). Bei fünf Betrieben („1“, „2“,

„6“, „8“, „9“, „10“) war der mittlere PCA QBA-Score von Dimension 1 am Morgen signifikant höher als am Vormittag. Bei den Betrieben 8, 9 und 10 war die Nachmittagsbeurteilung nicht signifikant von der Morgenbeurteilung unterschiedlich, wohingegen beim Betrieb 1 die Nachmittagsbeurteilung nicht signifikant von der Vormittagsbeurteilung abwich. Beim Betrieb 2 war die Nachmittagsbeurteilung sowohl von der Morgen- als auch von der Vormittagsbeurteilung signifikant unterschiedlich und beim Betrieb 6 war der Nachmittag weder vom Morgen noch vom Vormittag signifikant verschieden. Bei zwei Betrieben „3“ und „5“ war die „Aktivität“ bei der Nachmittagsbeurteilung signifikant niedriger, jedoch gab es keinen signifikanten Unterschied zwischen Morgen- und Vormittagsbeurteilung (Tabelle 7).

**Tabelle 7: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 2 für 10 Betrieben beurteilt von 13 BeobachterInnen**

<b>Betrieb</b>	<b>Morgen: Mittelwert (<math>\pm</math> SD)</b>	<b>Vormittag: Mittelwert (<math>\pm</math> SD)</b>	<b>Nachmittag: Mittelwert (<math>\pm</math> SD)</b>
<b>1</b>	<b>0,65 (<math>\pm</math> 0,82) a</b>	<b>-0,53 (<math>\pm</math> 0,81) b</b>	<b>-0,74 (<math>\pm</math> 0,76) b</b>
<b>2</b>	<b>0,73 (<math>\pm</math> 0,64) a</b>	<b>-1,13 (<math>\pm</math> 0,75) b</b>	<b>-0,14 (<math>\pm</math> 0,92) c</b>
<b>3</b>	<b>0,92 (<math>\pm</math> 0,84) a</b>	<b>0,80 (<math>\pm</math> 0,88) a</b>	<b>-0,21 (<math>\pm</math> 0,62) b</b>
<b>4</b>	-0,66 ( $\pm$ 0,64)	-0,87 ( $\pm$ 0,77)	-0,80 ( $\pm$ 0,73)
<b>5</b>	<b>0,82 (<math>\pm</math> 0,73) a</b>	<b>0,52 (<math>\pm</math> 0,81) a</b>	<b>-0,80 (<math>\pm</math> 0,91) b</b>
<b>6</b>	<b>0,32 (<math>\pm</math> 0,78) a</b>	<b>-0,46 (<math>\pm</math> 0,55) b</b>	<b>-0,29 (<math>\pm</math> 0,75) ab</b>
<b>7</b>	-0,01 ( $\pm$ 0,66)	-0,71 ( $\pm$ 0,72)	-0,27 ( $\pm$ 0,78)
<b>8</b>	<b>1,42 (<math>\pm</math> 0,72) a</b>	<b>-0,27 (<math>\pm</math> 0,74) b</b>	<b>1,01 (<math>\pm</math> 1,11) a</b>
<b>9</b>	<b>0,57 (<math>\pm</math> 0,63) a</b>	<b>-0,82 (<math>\pm</math> 0,67) b</b>	<b>0,64 (<math>\pm</math> 0,81) a</b>
<b>10</b>	<b>0,39 (<math>\pm</math> 0,66) a</b>	<b>-0,45 (<math>\pm</math> 0,53) b</b>	<b>0,36 (<math>\pm</math> 0,74) a</b>

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).

## 4.2.2 Reduzierter Datensatz

Im folgenden Abschnitt werden die Ergebnisse der WQ QBA-Scores und PCA QBA-Scores für Dimension 1 und 2, die für die unterschiedlichen Tageszeiten (Morgen, Vormittag und Nachmittag) berechnet wurden, dargestellt. Alle in diesem Abschnitt dargestellten Ergebnisse beziehen sich auf den *reduzierten Datensatz* der BeobachterInnengruppe, die 5 Personen umfasste.

### WQ QBA-Score

Mittels Varianzanalyse mit Messwiederholung wurde ein signifikanter Effekt der „Tageszeit“ ( $F_{2,80} = 17,923$ ;  $P < 0,001$ ) sowie der Interaktion zwischen „Tageszeit“ und „Betrieb“ gefunden ( $F_{18,80} = 3,363$ ;  $P < 0,001$ ).

In der einfaktoriellen ANOVA konnte für das Ergebnis der WQ QBA-Scores bei zwei Betrieben ein signifikanter Unterschied der Bewertung in Abhängigkeit von der Tageszeit gefunden werden. Bei Betrieb 2 und 6 wurden in der Morgenbewertung signifikant niedrigere Werte erreicht als in der Vormittagsbewertung, wohingegen sich die Werte am Morgen und am Nachmittag nicht signifikant voneinander unterschieden (Tabelle 8).

**Tabelle 8: Mittelwert ( $\pm$  SD) WQ QBA-Score für 10 Betriebe beurteilt von 5 BeobachterInnen**

Betrieb	Morgen: Mittelwert ( $\pm$ SD)	Vormittag: Mittelwert ( $\pm$ SD)	Nachmittag: Mittelwert ( $\pm$ SD)
1	52,3 ( $\pm$ 16,7)	38,5 ( $\pm$ 29,9)	27,7 ( $\pm$ 16,2)
<b>2</b>	<b>9,1 (<math>\pm</math> 16,0) a</b>	<b>39,7 (<math>\pm</math> 14,8) b</b>	<b>29,0 (<math>\pm</math> 18,6) ab</b>
3	31,9 ( $\pm$ 21,9)	49,4 ( $\pm$ 17,0)	40,5 ( $\pm$ 21,3)
4	16,0 ( $\pm$ 13,4)	23,1 ( $\pm$ 13,9)	17,8 ( $\pm$ 14,7)
5	37,2 ( $\pm$ 26,2)	30,5 ( $\pm$ 5,7)	22,7 ( $\pm$ 33,7)
<b>6</b>	<b>13,9 (<math>\pm</math> 13,1) a</b>	<b>53,6 (<math>\pm</math> 18,0) b</b>	<b>33,7 (<math>\pm</math> 17,5) ab</b>

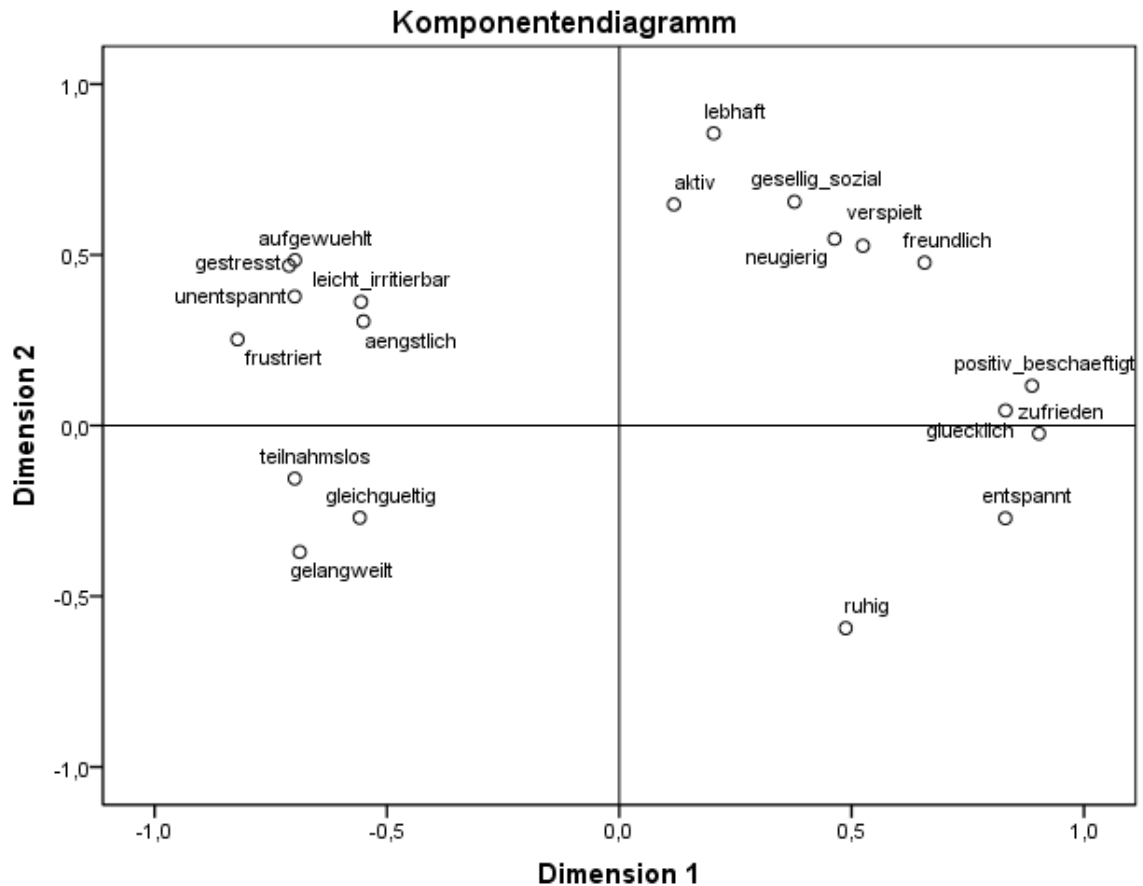
7	30,7 (± 25,4)	30,9 (± 17,7)	34,5 (± 27,4)
8	15,5 (± 16,6)	36,5 (± 28,0)	7,6 (± 12,7)
9	11,7 (± 15,6)	38,0 (± 19,6)	33,0 (± 29,4)
10	30,4 (± 25,7)	57,6 (± 19,9)	51,3 (± 23,1)

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).

### *Hauptkomponentenanalyse*

Der Barlett-Test auf Sphärizität ( $P < 0,001$ ) ergab, dass die Stichprobe für die Durchführung einer Hauptkomponentenanalyse geeignet war. Insgesamt drei Dimensionen erklärten jeweils mehr als 10% der Varianz. Dimension 1 erklärte 42% der Varianz, Dimension 2 erklärte 20% der Varianz. Bezieht man zusätzlich noch Dimension 3 ein, wurden 71% der Varianz erklärt.

Abbildung 2 enthält das Komponentendiagramm der PCA für die Dimensionen 1 und 2. Wie im Komponentendiagramm ersichtlich ist, laden auf Dimension 1 besonders stark die Begriffe 'unentspannt', 'frustriert' und 'gestresst' im negativen Bereich der Achse, sowie 'glücklich', 'zufrieden', 'positiv beschäftigt' im positiven Bereich. Dimension 1 kann mit dem Überbegriff „Stimmung“ interpretiert werden. Die Begriffe 'ruhig', 'gleichgültig', 'gelangweilt' laden stark im negativen Bereich von Dimension 2, sowie 'lebhaft', 'aktiv', 'gesellig/sozial' stark im positiven Bereich dieser Achse. Somit kann Dimension 2 mit dem Begriff „allgemeine Aktivität“ zusammengefasst werden.



**Abbildung 2: Verteilung der Begriffe im Komponentendiagramm.**

**Komponentenladung gibt die Korrelation der Begriffe mit den Dimensionen 1 und 2 an. Betriebe beurteilt durch 5 BeobachterInnen.**

### *PCA QBA-Score für Dimension 1*

Für den PCA QBA-Score für Dimension 1 lag ein signifikanter Effekt der „Tageszeit“ ( $F_{2,80} = 16,435$ ;  $P < 0,001$ ), sowie der Interaktion zwischen „Tageszeit“ und „Betrieb“ vor ( $F_{18,80} = 3,234$ ;  $P < 0,001$ ).

In der einfaktoriellen ANOVA wurde für die PCA QBA-Scores von Dimension 1 bei zwei Betrieben ein signifikanter Unterschied der Bewertung in Abhängigkeit von der Tageszeit gefunden. In Tabelle 9 sind die Mittelwerte der PCA QBA-Scores für Dimension 1 aus der Beurteilung von fünf BeobachterInnen dargestellt. Sowohl bei Betrieb 6 als auch bei Betrieb 9 waren die Werte für den PCA QBA-Score der Dimension 1 am Morgen signifikant niedriger als am Vormittag. Bei Betrieb 6 wich

das Ergebnis der Nachmittagsbewertung nicht signifikant von der Morgen- bzw. Vormittagsbewertung ab, wohingegen bei Betrieb 9 am Nachmittag signifikant niedrigere Werte erreicht wurden als bei der Vormittagsbewertung. Die Werte der Morgen- und Vormittagsbewertung waren bei diesem Betrieb nicht signifikant unterschiedlich.

**Tabelle 9: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 1 für 10 Betriebe beurteilt von 5 BeobachterInnen**

Betrieb	Morgen: Mittelwert ( $\pm$ SD)	Vormittag: Mittelwert ( $\pm$ SD)	Nachmittag: Mittelwert ( $\pm$ SD)
1	0,92 ( $\pm$ 0,60)	0,29 ( $\pm$ 1,26)	-0,93 ( $\pm$ 0,78)
2	-1,05 ( $\pm$ 1,01)	0,32 ( $\pm$ 0,63)	-0,14 ( $\pm$ 0,71)
3	0,11( $\pm$ 0,86)	0,78 ( $\pm$ 0,60)	0,33 ( $\pm$ 1,00)
4	-0,67( $\pm$ 0,64)	-0,41 ( $\pm$ 0,76)	-0,61 ( $\pm$ 0,78)
5	0,28 ( $\pm$ 1,01)	0,37 ( $\pm$ 0,18)	-0,58 ( $\pm$ 1,63)
<b>6</b>	<b>-0,77 (<math>\pm</math> 0,68) a</b>	<b>0,91 (<math>\pm</math> 0,66) b</b>	<b>0,22 (<math>\pm</math> 0,64) ab</b>
7	-0,01 ( $\pm$ 1,01)	0,13 ( $\pm$ 0,74)	0,08 ( $\pm$ 1,12)
8	-0,70 ( $\pm$ 1,01)	0,19 ( $\pm$ 1,01)	-1,11 ( $\pm$ 0,74)
<b>9</b>	<b>-0,88 (<math>\pm</math> 0,80) a</b>	<b>0,30 (<math>\pm</math> 0,70) b</b>	<b>0,18 (<math>\pm</math> 1,26) a</b>
10	0,02( $\pm$ 1,19)	1,01 ( $\pm$ 0,75)	0,84 ( $\pm$ 0,97)

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).

### PCA QBA-Score für Dimension 2

Für den PCA QBA-Score für Dimension 2 lag ein signifikanter Effekt der „Tageszeit“ ( $F_{2,80} = 43,090$ ;  $P < 0,001$ ) und der Interaktion von „Tageszeit“ und „Betrieb“ ( $F_{18,80} = 5,731$ ;  $P < 0,001$ ) vor.

Die einfaktorielle ANOVA ergab für die PCA QBA-Scores für Dimension 2 bei insgesamt drei Betrieben einen signifikanten Unterschied der Bewertung in Abhängigkeit von der Tageszeit. Bei Betrieb 2 waren die Werte für den PCA QBA-Score von Dimension 2 am Morgen signifikant höher als am Vormittag, wohingegen die Werte der Nachmittagsbewertung nicht signifikant unterschiedlich von den Werten der Morgen- und Vormittagsbewertung waren. Bei den Betrieben 5 und 9 waren die Werte der Morgen- und Vormittagsbewertung nicht signifikant verschieden, jedoch war bei Betrieb 5 der Wert der Nachmittagsbewertung signifikant niedriger als der Wert der Morgenbewertung und bei Betrieb 9 ergab die Nachmittagsbewertung signifikant höhere Werte als die Vormittagsbewertung (Tabelle 10).

**Tabelle 10: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 2 für 10 Betrieben beurteilt von 5 BeobachterInnen**

<b>Betrieb</b>	<b>Morgen: Mittelwert (<math>\pm</math> SD)</b>	<b>Vormittag: Mittelwert (<math>\pm</math> SD)</b>	<b>Nachmittag: Mittelwert (<math>\pm</math> SD)</b>
1	0,90 ( $\pm$ 1,01)	-0,51 ( $\pm$ 0,75)	-0,45 ( $\pm$ 0,66)
<b>2</b>	<b>0,73 (<math>\pm</math> 0,84) a</b>	<b>-1,39 (<math>\pm</math> 0,53) b</b>	<b>-1,05 (<math>\pm</math> 1,01) ab</b>
3	0,73 ( $\pm$ 1,01)	0,77 ( $\pm$ 1,13)	-0,30 ( $\pm$ 0,42)
4	-0,51 ( $\pm$ 0,75)	-0,93 ( $\pm$ 0,48)	-0,60 ( $\pm$ 0,45)
<b>5</b>	<b>0,90 (<math>\pm</math> 1,01) a</b>	<b>0,18 (<math>\pm</math> 0,62) ab</b>	<b>-0,80 (<math>\pm</math> 0,74) b</b>
6	-0,05 ( $\pm$ 0,71)	-0,08 ( $\pm$ 0,71)	-0,33 ( $\pm$ 0,67)
7	-0,07 ( $\pm$ 0,87)	-0,54 ( $\pm$ 0,55)	-0,37 ( $\pm$ 0,75)
8	1,31 ( $\pm$ 0,91)	-0,29 ( $\pm$ 0,78)	0,96 ( $\pm$ 1,55)
<b>9</b>	<b>0,57 (<math>\pm</math> 0,88) ab</b>	<b>-0,72 (<math>\pm</math> 0,61) a</b>	<b>0,80 (<math>\pm</math> 0,98) b</b>

---

10	0,41 ( $\pm 1,01$ )	-0,25 ( $\pm 0,71$ )	0,33 ( $\pm 0,82$ )
----	---------------------	----------------------	---------------------

---

Unterschiedliche Buchstaben innerhalb von Reihen bezeichnen signifikante Unterschiede ( $p < 0.05$ ).



## **5 Diskussion**

### **5.1 Methodik**

#### **5.1.1 Versuchsdesign**

##### *Anzahl der Betriebe*

Die Untersuchungen wurden auf zehn Betrieben durchgeführt. Zwar wäre eine größere Stichprobe wünschenswert gewesen, doch aufgrund der Tatsache, dass auch bei diesem moderaten Stichprobenumfang signifikante Effekte gefunden wurden, kann diese Anzahl, zur Beantwortung der vorliegenden Forschungsfragen, als ausreichend betrachtet werden.

##### *Variabilität der Untersuchungsbetriebe*

Bezüglich ihrer Strukturierung (Anzahl der Milchkühe, Haltungssystem) waren die Untersuchungsbetriebe recht einheitlich. Dies könnte, aufgrund der engen Bandbreite an Referenzvideos, die Beobachterübereinstimmung negativ beeinflusst haben. Dennoch waren die Scores über nahezu die gesamte WQ QBA-Punkteskala verteilt und reichten von stark negativen bis stark positiven PCA QBA-Scores, wodurch angezeigt wird, dass ein Beobachten des emotionalen Ausdrucks der Tiere ausreichte, um Betriebe voneinander zu unterscheiden.

Da es sich um Praxisbetriebe handelte, kam es bei einigen Einflussfaktoren zu gewissen Abweichungen. Diese betrafen die Fütterungs- und Melkzeiten. Am Morgen wurde mit dem Qualitative Behaviour Assessment erst begonnen, nachdem die Tiere aus der Fixierung im Fressgitter gelöst worden waren und sich wieder frei im Laufstall bewegen konnten und sich außer den an der Untersuchung beteiligten BeobachterInnen keine weiteren Personen im Stall befanden. Dieser Zeitpunkt variierte auf den Betrieben zwischen 7.00 Uhr und 8.45 Uhr. Bei einem Betrieb (Betrieb 1) konnte die Morgenbeobachtung erst um 9.30 Uhr beginnen, da eine frühere Beobachtung durch einen länger andauernden Fütterungsvorgang unterbrochen wurde. Die Vormittagsbeobachtung wurde zwischen 10.00 Uhr und 11.30 Uhr durchgeführt und die

Nachmittagsbeobachtungen erstreckten sich von 12.15 Uhr – 13.45 Uhr. Diese Zeitspannen ergeben sich einerseits daraus, dass zwischen den Beobachtungen ein zeitlicher Abstand von mindestens 90 Minuten eingehalten werden sollte und zum anderen die Untersuchung den auf den Betrieben praktizierten zeitlichen Routineablauf nicht beeinflussen durfte.

Abweichungen sind insbesondere im unterschiedlichen Management der Betriebe zu finden. Bei einem Betrieb (Betrieb 6) wurden die Tiere mittels automatischer Lockfütterung (Modell Lockboy, Fa. Wasserbauer), bei der kleine Kraffuttermengen angeboten werden, immer wieder zum Fressen an den Barren gelockt. Die restlichen Betriebe praktizierten ein mehrmals tägliches händisches bzw. maschinelles Heranschieben von weggeschobenem Futter an den Barren. Die Zeitpunkte und Häufigkeiten dieses Arbeitsgangs variierten von Betrieb zu Betrieb und stellen einen möglichen Einflussfaktor bezüglich der emotionale Befindlichkeit dar.

#### *Einfluss durch betriebsfremde Beobachter*

Die Wahl der drei Beobachtungszeitpunkte, morgens, vormittags und früher Nachmittag, ist mit der Durchführbarkeit im Rahmen des Welfare Quality®-Protokolls zu begründen. Ein praxisnaher Ansatz erschien wichtig und stärkt die Aussagekraft dieser Arbeit.

Einen späteren Zeitpunkt zu erheben wäre nicht sinnvoll gewesen, da für diesen Zeitpunkt gemäß Welfare Quality®-Protokoll die klinische Beurteilung vorgesehen ist und die Tiere sehr stark in ihrem natürlichen Verhalten gestört sind. Es war jedoch ausdrückliches Ziel, sich an der Vorgehensweise des Welfare Quality®-Protokolls zu orientieren.

Bei den gewählten Beobachtungszeiten wurde versucht, den Einfluss der Beobachter so gering wie möglich zu halten. Zwischen Eintreffen der Beobachter am Untersuchungsbetrieb bis zur ersten Beobachtung wurde ein zeitlicher Abstand von mindestens 30 Minuten eingehalten, damit sich die Tiere an die veränderte Situation gewöhnen konnten. Dass die Tiere während der Beobachtung nur geringes Interesse an den Beobachtern zeigten, kann als Gewöhnung interpretiert werden.

Während der 20-minütigen Beobachtung verhielten sich die Beobachter ruhig und versuchten, die Tiere in keinerlei Weise zu stören oder zu beeinflussen. Nichtsdestotrotz muss in Betracht gezogen werden, dass auch gut an den Menschen gewöhnte Tiere durch dessen Anwesenheit in ihrem Verhalten beeinflusst sein können (Martin und Bateson, 2007, 31).

### **5.1.2 Vorteile, Nachteile und Grenzen der Videobewertung**

Um den Einfluss der Tageszeit auf die Beurteilung mittels Qualitative Behaviour Assessment zu prüfen, ist die Verwendung von Videotechnik unerlässlich. Dabei ist darauf zu achten, dass die verwendete Technik an die jeweiligen Untersuchungstiere, das Haltungssystem sowie die Fragestellung angepasst ist.

Ein Vorteil der Videobewertung ist es, dass die BeobachterInnen ihre Bewertung unbeeinflusst von der aktuellen Tageszeit durchführen können. Dazu wurden die Videoclips in durchmischter Reihenfolge den BeobachterInnen gezeigt. Auch konnte durch die Videobeobachtung eine höhere Zahl an Beobachterinnen und Beobachtern erreicht werden, als dies aus organisatorischen und zeitlichen Gründen bei einer On-farm-Beurteilung möglich gewesen wäre.

Ein eindeutiger Nachteil der Videobewertung ist die Zweidimensionalität der Aufnahmen. Die Beurteilung des Tierverhaltens kann dadurch erschwert sein. Die geringe Auflösung, aufgrund von unzureichenden Lichtverhältnissen im Stall, das eingeschränkte Sichtfeld und die Tiefenschärfe reichen nicht immer aus, um die erforderlichen Daten aus dem Filmmaterial zu gewinnen (Martin und Bateson, 2007, 101f).

Da bei der vorliegenden Arbeit das Verhalten von Milchkühen auf Herdenniveau beurteilt wurde und ein Filmen meist nur vom Futtertisch aus möglich war, musste die Kamera über eine ausreichende Weitwinkelfunktion verfügen. Dennoch muss zum Filmen eine möglichst weit vom Barren entfernte Position gewählt werden, damit ein möglichst großer Stallausschnitt am Video abgebildet werden kann. Dies ist dann schwierig, wenn sich auch auf der anderen Seite des Futtertisches Boxen befinden, weil die sich darin befindlichen Tiere äußerst gerne das aufgestellte

Stativ erkunden. Teilweise kann es schwierig sein, Tiere, die sich weiter hinten im Stall befinden, ordentlich zu beurteilen. Die Videobeobachtung muss dann von den BeobachterInnen sehr diszipliniert durchgeführt werden, indem auch das Geschehen im Hintergrund in die Bewertung miteinbezogen wird. Dies kann grundsätzlich als eine Herausforderung der Herdenbeurteilung betrachtet werden. Manchmal passiert es, dass sich Tiere so vor der Kamera platzieren, dass andere Tiere verdeckt werden. Bei der Direktbeobachtung kann dies häufig durch einen kleinen Positionswechsel ausgeglichen werden, was bei der Videobeobachtung nicht möglich ist (Cimer, 2011).

Die vorliegende Arbeit begegnete dieser Herausforderung damit, dass Daten der Videobeurteilung nicht mit denen der Direktbeurteilung verglichen wurden. Außerdem wurden die Aufnahmen zu den unterschiedlichen Tageszeiten alle von derselben Position aus gemacht. Für die gekürzten Videoclips schließlich wurden im Tagesverlauf immer dieselben Stallausschnitte gewählt. Ein ebenfalls wichtiger Punkt bei der Videobeurteilung ist der Ton. Dadurch wird die Atmosphäre im Stall spürbar. Aus diesem Grund wurden alle Videos mit Ton gezeigt.

Ein großer Vorteil der Videobeurteilung für die Fragestellung der vorliegenden Arbeit ist, dass die gezeigten Videoclips hinsichtlich der Tageszeit randomisiert werden können, sodass die BeobachterInnen unvoreingenommen beurteilen können. Ein weiterer Vorteil besteht darin, dass es möglich ist, die Beurteilung von einer größeren Zahl an BeobachterInnen durchführen zu lassen, als dies aus organisatorischen Gründen bei einer On-farm-Beurteilung möglich wäre. Das Videomaterial kann mehrmals beurteilt werden, wodurch zusätzliche Daten zur Validierung der Methode gewonnen werden können. Außerdem ist das Beobachtungsmaterial „konserviert“ und steht für weitere Untersuchungen zur Verfügung. Es sollte ebenfalls erwähnt werden, dass die Anschaffung einer entsprechenden Videoausrüstung sowie eines ausreichend großen Speichermediums mit erheblichen Kosten verbunden ist und die Vorbereitung des Bildmaterials für die Videobeobachtung viel Zeit in Anspruch nimmt.

## **5.2 Ergebnisse**

Das Ziel der vorliegenden Arbeit war es, zu testen, ob es hinsichtlich der emotionalen Befindlichkeit bei Milchkühen bei der Beurteilung mittels Qualitativer Verhaltensbeurteilung in Abhängigkeit von der Tageszeit zu unterschiedlichen Ergebnissen kommt. Dies ist die erste derartige Überprüfung und stellt somit eine Erweiterung von vorhergehenden Studien zur Überprüfung der Zuverlässigkeit von QBA dar (Andreasen et al., 2013; Bokkers et al., 2012; Rousing und Wemelsfelder, 2006; Wemelsfelder et al., 2009d).

### **5.2.1 BeobachterInnenübereinstimmung beim Qualitative Behaviour Assessment**

Die Übereinstimmung zwischen BeobachterInnen ist neben der Übereinstimmung innerhalb von BeobachterInnen einer jener Werte, die eine Aussage darüber zulassen, ob eine Beurteilungsmethode als wissenschaftlich valide gelten kann (Meagher, 2009). Die Übereinstimmung zwischen BeobachterInnen für die QBA-Ergebnisse in dieser Studie waren gering bis mäßig bei der 13 Personen umfassenden Beobachtergruppe und mäßig bis hoch im reduzierten Datensatz mit fünf BeobachterInnen. Ein Koeffizient für Kendall's  $W$  von 0,7, der als Grenzwert für einen akzeptablen Wert des Korrelationskoeffizienten in der Beobachterübereinstimmung gilt (Martin und Bateson, 2007, 51ff), wurde nur im reduzierten Datensatz mit ausgewählten BeobachterInnen gefunden. Die berechneten Werte entsprechen jedoch in der Größenordnung den Angaben von Wemelsfelder et al. (2009c) und Bokkers et al. (2012), die ebenfalls nur geringe Beobachterübereinstimmungen in der QBA-Beurteilung von Milchkühen fanden. Auch bei der Entwicklung der Fixed Terms List für QBA bei Milchkühen (Wemelsfelder et al., 2009c) wurde, wie in der vorliegenden Arbeit, eine Hauptkomponentenanalyse verwendet und die Übereinstimmung zwischen BeobachterInnen mittels Kendall's Übereinstimmungskoeffizient  $W$  ermittelt. Kendall's  $W$  erreichte in dieser Studie ähnlich geringe Werte wie in der vorliegenden Arbeit: Der Kendall's  $W$  erreichte für Dimension 1 einen Wert von 0,38 und für Dimension 2 von 0,46, wenn in die Bewertung 20 vorgegebene

Begriffe einbezogen wurden. In der genannten Studie konnten die BeobachterInnen ursprünglich die Bewertung mit 28 vorgegebenen Begriffen durchführen, was zu etwas höheren Kendall's *W*-Werten führte (Wemelsfelder et al., 2009c).

Bei den BeobachterInnen, die an der Videobeurteilung für die vorliegende Arbeit teilnahmen, handelte es sich um Studierende und Nutztierwissenschaftlerinnen. Daher gab es zwischen den BeobachterInnen Unterschiede in den Kenntnissen über das Verhalten von Milchkühen. Außerdem waren sechs BeobachterInnen bereits erfahren in der Beurteilung von Milchkuhverhalten mittels Qualitative Behaviour Assessment, wohingegen die übrigen sieben BeobachterInnen diesbezüglich Laien waren. Doch bereits Wemelsfelder und Lawrence (2001b) stellten fest, dass der fachliche Hintergrund des Beobachters bzw. der Beobachterin das Ergebnis nicht übermäßig beeinflusst.

Dies legt nahe, den Grund für die geringe Beobachterübereinstimmung an anderer Stelle zu suchen. Ein Punkt, die Anzahl an Begriffen, wurde bereits genannt. Von einigen BeobachterInnen wurde rückgemeldet, dass sie Schwierigkeiten hatten, einzelne Begriffe zu beurteilen, beziehungsweise dass für bestimmte Verhaltensweisen keine Begriffe zur Verfügung stehen. Als schwierig zu beurteilen galten die Begriffe „unentspannt“ und „glücklich“. Eine Überarbeitung der vorgegebenen Begriffe könnte möglicherweise zu einer einheitlicheren Beurteilung führen.

Bereits Wemelsfelder et al. (2009c) haben aufgrund geringer Beobachterübereinstimmung in ihrer oben erwähnten Studie festgestellt, dass es wichtig ist, diese bereits im Training zu überprüfen, entweder über Videobeurteilung oder besser noch Livebeobachtung. Dieses Training soll so lange durchgeführt werden, bis eine ausreichend hohe Übereinstimmung in der Beurteilung erreicht ist. Sollten Probleme auftauchen, schlagen die ForscherInnen vor, ein zusätzliches Training in der Beobachtung und Interpretation von Rinderverhalten vorzunehmen.

Das Training, welches der Videobeobachtung der vorliegenden Studie vorausgegangen ist, beinhaltete einerseits eine Einführung in die Methode des Qualitative Behaviour Assessment sowie eine praktische Anleitung zur Beurteilung mittels Fixed Terms. Da es wichtig ist, dass jede Beobachterin und jeder Beobachter die Begriffe gleich interpretiert (Meagher, 2009), wurde anschließend

jeder Begriff einzeln diskutiert. Außerdem wurde ein Videoclip zur Probe bewertet und die Bewertung gemeinsam besprochen. Insgesamt nahm das Training 90 Minuten in Anspruch. Eine weitere Beurteilung von Videoclips zu Trainingszwecken wäre sicherlich wünschenswert gewesen, hätte jedoch den zeitlichen Rahmen der Videobeurteilung überschritten.

Die Ergebnisse aus dem reduzierten Datensatz weisen darauf hin, dass eine akzeptable Beobachterübereinstimmung durchaus möglich ist. Allerdings wurden in der wiederholten Messung trotz der zum Teil geringen Beobachterübereinstimmung sowohl im vollständigen Datensatz der 13 BeobachterInnen als auch im reduzierten Datensatz der fünf BeobachterInnen weitestgehend die gleichen Effekte gefunden. Obwohl es natürlich für eine wirklich zuverlässige Messung, wenn zum Beispiel Betriebe dadurch zertifiziert werden, notwendig ist, eine ausreichend hohe Beobachterübereinstimmung zu erreichen, ist die Beobachterübereinstimmung für die in dieser Arbeit gestellte Forschungsfrage von untergeordneter Bedeutung. Denn das Interesse galt vordergründig den tageszeitlichen Unterschieden innerhalb eines Betriebes und nicht der absoluten Beurteilungsübereinstimmung. Indem der Datensatz auf fünf BeobachterInnen mit mäßiger bis hoher Beobachterübereinstimmung reduziert wurde, konnte gezeigt werden, dass es dabei zu keinem Unterschied im Tageszeiteffekt, verglichen mit dem vollständigen Datensatz kam. Daher konzentriert sich der nun folgende Abschnitt der Diskussion auf die gefundenen Haupteffekte des vollständigen Datensatzes der 13 BeobachterInnen.

### **5.2.2 Einfluss der Tageszeit auf die qualitative Verhaltensbeurteilung**

In der vorliegenden Studie wurden zwei unterschiedliche Ansätze der Datenanalyse angewendet: Der WQ QBA-Score und die PCA QBA-Scores. Der WQ QBA-Score basiert auf vorgegebenen Koeffizienten bzw. Gewichtungen der Bewertungsbegriffe, die aufgrund eines Referenzdatensatzes berechnet wurden. Wir nahmen an, dass die Hauptkomponentenanalyse durch das exaktere Wiedergeben der Dimensionen, wie sie durch die BeobachterInnen definiert wurden, zu anderen Ergebnissen führen würde. Allerdings entspricht die erste Dimension der Hauptkomponentenanalyse ziemlich genau der WQ-Dimension:

Wie in WQ, so unterscheidet auch die erste Dimension der Hauptkomponentenanalyse, die den größten Teil der Variation erklärt, zwischen positiver und negativer Stimmung und hat somit unmittelbare Bedeutung für die Erhebung des Welfare Kriteriums „Positive emotionale Befindlichkeit“.

Dimension 2 unterscheidet zwischen hoher und geringer Aktivität, was keine unmittelbare Relevanz für das Wohlergehen hat (Wemelsfelder et al., 2009c) und deshalb als weniger bedeutsam für die Beurteilung betrachtet werden kann. Die WQ-Definition der „Stimmungsdimension“ konnte durch die Hauptkomponentenanalyse bestätigt werden, was darauf hinweist, dass die QBA-Analyse gemäß WQ dazu geeignet ist, die Gewichtung der Begriffe durch die BeobachterInnen entsprechend wiederzugeben.

Die Varianzanalyse mit Messwiederholungen ergab ähnliche Effekte des Haupteffekts und des Interaktionseffekts für 'Tageszeit' und 'Betrieb' für beide Analysemethoden, d.h. WQ QBA-Score und die PCA QBA-Scores. Für die Scores die in Zusammenhang mit „emotionaler Befindlichkeit“/ 'Stimmung' stehen, wurden drei Betriebe mit signifikanten Unterschieden in Abhängigkeit von der Beobachtungszeit identifiziert, wobei zwei der drei Betriebe in beiden Analysemethoden dieselben waren und die Scores über die Beobachtungszeitpunkte hinweg innerhalb der Betriebe jeweils dem selben Muster folgten.

Da die Bedeutung von WQ und Dimension 1 der Hauptkomponentenanalyse sowie das Muster der tageszeitlichen Unterschiede sehr ähnlich sind, können die Unterschiede in den Betrieben, die im Tagesverlauf signifikante Unterschiede aufwiesen, eher durch die hohe Varianz der Scores innerhalb eines Betriebes als durch qualitative Unterschiede erklärt werden.

Die PCA QBA-Scores für Dimension 2 ('Aktivität') unterschieden sich signifikant voneinander in acht von zehn Betrieben. Außer einem 'Aktivitäts'-Hoch in allen Betrieben am Morgen (signifikant oder nicht), konnte kein durchgängiges Muster erkannt werden. Im Gegensatz zur 'Stimmung' weist Aktivität nicht explizit auf das Wohlergehen hin. Jedoch könnten die unterschiedlichen Muster der Aktivitäts-Scores auf den einzelnen Betrieben den Einfluss des Betriebsmanagements widerspiegeln (DeVries und von Keyserlingk, 2005).

Während beim ersten Beobachtungszeitpunkt am frühen Morgen, durch die Futtervorlage gleich nach dem Morgenmelken, die Situation für die Tiere auf den



verschiedenen Betrieben wahrscheinlich ähnlich ist, könnte sich die Situation ein paar Stunden später sehr viel mehr unterscheiden (z.B. hinsichtlich Futtermalage, anderer täglicher Routinehandlungen oder tierärztlicher Behandlungen), und dementsprechend könnte auch die emotionale Befindlichkeit davon beeinflusst sein. Da die Hauptkomponentenanalyse orthogonale Dimensionen erzeugt, wurden durch die BeobachterInnen 'Stimmung' und 'Aktivität' unabhängig voneinander beurteilt. So konnten Tiere zum Beispiel als „aktiv“ und „in guter Stimmung“, als auch als „ruhig“ und in „schlechter Stimmung“ wahrgenommen werden. Dennoch könnten die unter den Betrieben variierenden Aktivitätsmuster das Auftreten verschiedener Ereignisse anzeigen, wodurch wiederum die Stimmung der Tiere beeinflusst sein könnte. Allerdings konnte keine Kontinuität im Muster der „emotionalen Befindlichkeit“/ 'Stimmung' gefunden werden. Nur ein einziger Betrieb erzielte die höchste Bewertung der „emotionalen Befindlichkeit“/'Stimmung' am frühen Morgen, verglichen mit den anderen Bewertungszeitpunkten (Betrieb 1 in der Bewertung mittels WQ QBA-Score und PCA QBA-Score für Dimension 1). Dementsprechend könnte eine bloß einmalige, am frühen Morgen durchgeführte QBA Bewertung, die durchschnittliche Stimmung am Betrieb häufig unterschätzen. Jedoch könnte andererseits eine Bewertung zu diesem Zeitpunkt die am stärksten standardisierte Situation über Betriebe hinweg bedeuten.

Huzzey et al. (2006) weisen darauf hin, dass durch eine hohe Besatzdichte die Konkurrenz am Fressplatz zunimmt. Die ForscherInnen beobachteten dabei mehr Verdrängungen besonders von niederrangigen Kühen. Im Vergleich dazu fraßen die Tiere länger und standen weniger inaktiv herum, wenn die Besatzdichte niedriger war. Bei bestimmten Betrieben könnten Restriktionen in der Haltung, besonders am Morgen, dem Zeitpunkt des umfangreichsten Fressens (DeVries et al., 2003), zu vermehrten agonistischen Interaktionen, wie Verdrängungen und Kopfstößen geführt haben. Diese agonistischen Interaktionen könnten sich negativ auf die Bewertung der „emotionale Befindlichkeit“/ 'Stimmung' ausgewirkt haben. Möglicherweise stellen sie eine Erklärung für die zum Teil niedrigen Scores der Morgenbewertung dar.

Das Ziel des WQ-Protokolls ist es, ein standardisiertes Beurteilungssystem für die On-Farm-Bewertung von Milchviehbetrieben zur Verfügung zu haben (Blokhuis et

al., 2008). WQ definierte einen zeitlichen Ablaufplan für die Betriebsbesuche. Außerhalb vom Morgenmelken als Taktgeber folgt QBA der Messung der „Ausweichdistanz am Fressgitter“. Zwar gibt das WQ-Protokoll keine genaue Zeitangabe, wann QBA durchgeführt werden soll, doch in allen für diese Studie besuchten Milchviehbetrieben war das Morgenmelken zwischen 0700h und 0800h abgeschlossen. Daraus ergab sich der Zeitpunkt für die QBA-Bewertung mit  $\pm$  0800h. Ein Vorteil, QBA generell am Beginn der Erhebungen eines Betriebsbesuches durchzuführen ist, dass Beobachter bei der QBA Bewertung nicht durch Informationen, die durch bereits erfolgte Erhebungen gesammelt wurden, beeinflusst sind. Jedenfalls aber sollte bei der Interpretation der QBA Ergebnisse berücksichtigt werden, dass unterschiedliche Beobachtungszeitpunkte zu Variationen in den QBA Ergebnissen führen können. Eine wiederholte, über den Betriebsbesuch verteilte QBA-Beurteilung würde sehr wahrscheinlich die „wahre“ emotionale Befindlichkeit der Tiere auf manchen Betrieben besser wiedergeben, als eine einzige Beurteilung. Jedoch würde andererseits, durch das Erinnern an Ergebnisse von zuvor durchgeführten tierbezogenen Messungen, das Risiko einer voreingenommenen Beurteilung unweigerlich steigen. Eine einzige QBA-Bewertung könnte zu einer Fehleinschätzung der längerfristigen emotionalen Befindlichkeit der Tiere eines Betriebes führen, hingegen sind die BeobachterInnen unvoreingenommen und die Ergebnisse aufgrund der relativ einheitlichen Situation nach dem Morgenmelken besser miteinander vergleichbar. Auf Basis der gewonnenen Ergebnisse kann empfohlen werden, QBA zu unterschiedlichen Zeitpunkten über den Tag verteilt durchzuführen (z.B. morgens und vormittags), wenn eine repräsentative Bewertung der emotionalen Befindlichkeit von Milchkühen angestrebt wird. Hingegen lässt die alleinige QBA Beurteilung am Morgen, wie sie gemäß dem WQ Protokoll durchgeführt wird, nur eine Aussage über den emotionalen Zustand zu diesem bestimmten Zeitpunkt zu. Hinsichtlich experimenteller QBA-Studien sollte der Vergleichbarkeit der untersuchten Situation eine höhere Priorität gegeben werden, um optimale Voraussetzungen für eine akzeptable Wiederholbarkeit zu schaffen. In vorausgegangenen Studien, die zur Untersuchung der Zuverlässigkeit von QBA durchgeführt wurden, schwankten die Beobachtungszeitpunkte beträchtlich. In einer Studie von Wemelsfelder *et al.* (2009c) zur Entwicklung der Fixed Terms für QBA bei Milchkühen, wurden die Beurteilungen zwischen 1000h und 1300h

durchgeführt und eine nur geringe Beobachterübereinstimmung wurde erreicht (Kendall's  $W$  Koeffizient von 0,38). Andreasen *et al.* (2013) untersuchten das Potential von QBA als eigenständige Untersuchungsmethode, um Betriebe mit beeinträchtigtem Tierwohlergehen zu identifizieren. Dazu wurden Beziehungen zwischen QBA Scores und anderen WQ Messungen ermittelt. Die QBA-Beurteilungen in dieser Studie erfolgten von 0940h bis 1645h durch zwei beobachtende Personen, die ausschließlich die qualitative Beurteilung mit QBA durchführten, während das vollständige WQ-Protokoll für Milchkühe, das durch eine andere Person abgearbeitet wurde, zwischen 0415h und 0900h startete. Die Beobachterübereinstimmung zwischen den beiden Personen, die gleichzeitig QBA durchführten, war hoch signifikant, jedoch wurde keine signifikante Korrelation zwischen den QBA-Scores der Personen, die ausschließlich QBA durchführten, und der Person, die das vollständige WQ-Protokoll erhob, gefunden. Die AutorInnen argumentieren, dass verschiedene praktische Umstände beim Fehlen der Korrelation eine Rolle gespielt haben dürften, z.B. unterschiedlicher Trainingsstatus der Beurteiler, zu ähnliche Betriebe, oder Schwierigkeiten beim Interpretieren der Körpersprache der Kühe. Der Einfluss der Tageszeit auf die QBA-Ergebnisse wurde in keiner der beiden Studien diskutiert. Die Ergebnisse der vorliegenden Studie weisen darauf hin, dass geringe Beobachterübereinstimmung im Zeitablauf auf geringe Wiederholbarkeit der Ergebnisse zu verschiedenen Zeitpunkten zurückzuführen sein könnte, d.h. die erzielten Ergebnisse könnten bloß die Unterschiede in der beobachteten Situation widerspiegeln und nicht eine geringe Beobachterübereinstimmung.

## **6 Schlussfolgerung und Ausblick**

Die Ergebnisse dieser Studie zeigen, dass bei bestimmten Betrieben die QBA Ergebnisse in Abhängigkeit vom Beobachtungszeitpunkt signifikant variieren können. Dieser Effekt bestand sowohl für den WQ QBA-Score als auch für Dimension 1 des PCA QBA-Scores; beide spiegeln die emotionale Befindlichkeit der Tiere wider. Die Ergebnisse der vorliegenden Arbeit legen nahe, sorgfältig die Zielsetzung der qualitativen Verhaltensbeurteilung zu berücksichtigen: wenn Aussagen über die „durchschnittliche Stimmung“ gemacht werden sollen, sollte die QBA-Bewertung zu verschiedenen Tageszeiten durchgeführt werden, wohingegen einem standardisierten Beobachtungszeitpunkt, wie er zum Beispiel im WQ-Protokoll vorgeschlagen wird, Priorität gegeben werden sollte, wenn Vergleichbarkeit unbedingt erforderlich ist.

Die vorliegende Arbeit nahm die Herausforderung an, die Messung der emotionalen Befindlichkeit von Milchkühen unter Verwendung des WQ-Protokolls zu verbessern. Die erzielten Ergebnisse können dabei helfen, die Zuverlässigkeit von On-Farm-Bewertungen zu verbessern, um gültige Aussagen über das Wohlergehen der Tiere eines bestimmten Betriebs machen zu können. Das ist sowohl für das Vertrauen der Konsumenten als auch der Landwirte in die Bewertung wichtig und weitere Verbesserungen von On-Farm-Bewertungen können darauf aufbauen.

## 7 Zusammenfassung

Das Qualitative Behaviour Assessment, bei dem vorgegebene Begriffe verwendet werden, bildet einen Teilbereich des Welfare Quality® (WQ)-Bewertungsprotokolls für Milchkühe. Wie für alle Methoden zur Erfassung von Wohlergehen, hat auch für QBA die Zuverlässigkeit der Erhebungsmethode eine große Bedeutung. Dies gilt insbesondere dann, wenn damit Betriebe zertifiziert werden sollen. Die vorliegende Arbeit untersuchte die Wiederholbarkeit von QBA-Ergebnissen über drei verschiedene Beobachtungszeitpunkte während des Tagesverlaufs (Morgen, Vormittag und Nachmittag). Dazu wurden insgesamt 30 Videoclips á 4 min von zehn Milchviehbetrieben durch 13 BeobachterInnen mittels Beurteilungsbogen mit 20 Begriffen bewertet. QBA-Scores für das Kriterium 'Positive emotionale Befindlichkeit' wurden einerseits gemäß Welfare Quality® (WQ)-Bewertungsprotokoll berechnet und zusätzlich wurde eine Hauptkomponentenanalyse durchgeführt. Aus letzterer gingen zwei Hauptkomponenten hervor: Die Komponente, die mit 'Stimmung' beschrieben werden kann, entspricht dem WQ-Kriterium 'Positive emotionale Befindlichkeit', die andere Komponente ist 'Aktivität'. Unabhängig davon, ob die Ergebnisse mittels WQ bzw. Hauptkomponentenanalyse berechnet wurden, ergab die Varianzanalyse mit Messwiederholungen einen signifikanten Effekt der Tageszeit in Abhängigkeit vom Betrieb. Um die Zuverlässigkeit der Beurteilung von Wohlergehen bei Milchkühen zu verbessern, sollte beachtet werden, dass zumindest in einigen Betrieben die Beobachtungszeit das QBA-Ergebnis beeinflussen kann.

## **Summary**

*Qualitative Behaviour Assessment (QBA) using predefined descriptors forms part of the Welfare Quality® (WQ) assessment protocol for dairy cattle. As for other on-farm measures of welfare, reliability is an important issue especially if farms are to be certified. This study investigated the repeatability of QBA results across three different observation times during the day (early morning, late morning, early afternoon). For this purpose, in total 30 4-min video clips from ten commercial dairy farms were assessed by 13 observers using visual analogue scales for 20 terms. QBA scores for 'emotional state' were computed according to the Welfare Quality® protocol and additionally a Principal Component Analysis was carried out. The latter revealed two main dimensions which may be described by 'mood' thus corresponding to the 'emotional state' score of the WQ protocol and by 'activity'. Independent of type of parameter, i.e. scores derived according to WQ or from PCA, repeated measures ANOVA revealed a significant effect of observation time depending on the farm. In order to improve reliability of welfare assessment in dairy cattle it should be considered that at least in some farms the observation time can affect QBA results.*

## 8 Danksagung

Zuerst möchte ich mich bei meinen Eltern für die Unterstützung während meines Studiums bedanken. Durch ihre Geduld und die großzügigen finanziellen Beiträge haben sie mir ermöglicht, meinen vielfältigen Interessen nachgehen zu können, ohne den Druck, mein Studium möglichst bald abschließen zu müssen. Weiters gilt mein Dank meiner Frau Katharina Schwed, die mich in meiner Idee das Masterstudium „Ökologische Landwirtschaft“ zu beginnen sehr bestärkt hat und mir während meines Studiums immer mit Rat und Tat zur Seite gestanden ist. Nicht vergessen möchte ich an dieser Stelle meinen Schwiegervater Gerhard Schwed. Seine Kompetenz in allen EDV-Fragen hat mir sehr weitergeholfen. Allen Begleiterinnen und Begleitern während meines Studiums, vielen Dank für die gemeinsamen Stunden.

Für die Unterstützung bei der Durchführung meiner Masterarbeit bedanke ich mich herzlich bei meinem Betreuer Univ. Prof. Dr. Christoph Winckler für seine methodische Begleitung und konstruktiven Beiträge, DI Lukas Tremetsberger und Dipl. Biol. Anke Gutmann für die tatkräftige Unterstützung und ihre Bereitschaft, viel Zeit zu investieren, damit diese Arbeit gelingt. Danke an die Universities Federation for Animal Welfare für die finanzielle Unterstützung durch das UFAW Animal Welfare Student Scholarship 2012 und die Möglichkeit, die Ergebnisse meiner Arbeit bei internationalen Veranstaltungen vorzustellen. Ich möchte mich auch bei allen Betrieben bedanken, die an der Studie teilgenommen haben und uns warmen Tee gebracht haben, wenn Lukas und ich im Stall froren. Ein großes Dankeschön auch an die Beobachterinnen und Beobachter, die ihre wertvolle Zeit in diese Studie investiert haben. Zuletzt erweitere ich meine Wertschätzung auf alle Kühe, die maßgebend an dieser Arbeit beteiligt waren.

## 9 Literaturverzeichnis

- Andreasen, S.N., Wemelsfelder, F., Sandøe, 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, 9–17.
- Blokhuis, H.J., Jones, R.B., Geers, R., Miele, M., Veissier, I., 2003. Measuring and monitoring animal welfare: transparency in the food product quality chain. *Animal Welfare* 12, 445–455.
- Blokhuis, H.J., Keeling, L.J., Gavinelli, A., Serratos, J., 2008. Animal welfare's impact on the food chain. *Trends in food science and technology* 79–87.
- Bokkers, E., De Vries, M., Antonissen, I., De Boer, I., 2012. Inter- and intra-observer reliability of experienced and inexperienced observers for the Qualitative Behaviour Assessment in dairy cattle. *Animal Welfare* 307–318.
- Broom, D.M., 1986. Indicators of poor welfare. *British Veterinary Journal* 142, 524–526.
- Canali, E., Keeling, L., 2009. Welfare Quality® project: from scientific research to on farm assessment of animal welfare. *Italian Journal of Animal Science* 900–903.
- Cimer, K., 2011. Einfluss des Beobachters auf das Verhalten von Mastschweinen. Wien, Masterarbeit an der Universität für Bodenkultur.
- DeVries, T.J., von Keyserlingk, M.A.G., Beauchemin, K.A., 2003. Diurnal Feeding Pattern of Lactating Dairy Cows. *Journal of Dairy Science* 86, 4079–4082.
- DeVries, T.J., von Keyserlingk, M.A.G., 2005. Time of Feed Delivery Affects the Feeding and Lying Patterns of Dairy Cows. *Journal of Dairy Science* 88, 625–631.
- Fowler, J., Cohen, L., Jarvis, P., 1998. *Practical statistics for field biology*. Wiley, Chichester [u.a.].
- Fraser, D., 2003. Assessing animal welfare at the farm and group level: The interplay of science and values. *Animal Welfare* 12, 433–443.



- Harcourt, A.H., 1978. Activity periods and pattern of social interaction: a neglected problem. *Behaviour* 66, 121–135.
- Houpt, K.A., 2011. *Domestic Animal Behaviour for Veterinarians & Animal Scientists*. Wiley-Blackwell, Ames, Iowa.
- Huzzey, J.M., DeVries, T.J., Volois, P., von Keyserlingk, M.A.G., 2006. Stocking density and feed barrier design affect the feeding and social behavior of dairy cattle. *Journal of Dairy Science* 89, 126–133.
- Kirchner, M., 2012. Implementation of an animal-based welfare assessment system in beef bull farms. Universität für Bodenkultur, Wien.
- Knierim, U., Winckler, C., 2009. On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare* 18, 451–458.
- Main, D.C.J., Kent, J.P., Wemelsfelder, F., Ofner, E., Tuytens, F.A.M., 2003. Applications for methods of on-farm welfare assessment. *Animal Welfare* 12, 523–528.
- Martin, P., Bateson, P., 2007. *Measuring behaviour, An Introductory Guide*, Third Edition. Cambridge University Press, Cambridge.
- McGarigal, K., Cushman, S., Stafford, S.G., 2000. *Multivariate statistics for wildlife and ecology research*. Springer, New York.
- Meagher, R., 2009. Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science* 119, 1–14.
- Miele, M., Parisi, V., 2001. L'Etica del Mangiare, i valori e le preoccupazioni dei consumatori per il benessere animale negli allevamenti: un'applicazione dell'analisi Means-end Chain. *Rivista di Economia Agraria* 81–103.
- Minero, M., Tosi, M.V., Canali, E., Wemelsfelder, F., 2009. Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science* 116, 74–81.
- Müllner, B., 2011. Comparison of free choice profiling and fixed terms for qualitative behaviour assessment in dairy cattle. Wien, Masterarbeit an der Universität für Bodenkultur.

- Rousing, T., Wemelsfelder, F., 2006. Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science* 101, 40–53.
- Rutherford, K.M.D., Donald, R.D., Lawrence, A.B., Wemelsfelder, F., 2012. Qualitative Behavioural Assessment of emotionality in pigs. *Applied Animal Behaviour Science* 139, 218–224.
- Sachser, N., 2009. Das Wohlergehen der Tiere. In: Ach, J., Stephany, M. (Hrsg.), *Die Frage Nach Dem Tier. Interdisziplinäre Perspektiven Auf Das Mensch-Tier-Verhältnis*. LIT Verlag, Berlin.
- Spooler, H., De Rosa, G., Hörning, B., Waiblinger, S., Wemelsfelder, F., 2003. Integrating parameters to assess on-farm welfare. *Animal Welfare* 12, 529–534.
- Welfare Quality®, 2009a. Welfare Quality® Assessment Protocol for Poultry. Welfare Quality Consortium, Lelystad, The Netherlands.
- Welfare Quality®, 2009b. Welfare Quality® Assessment Protocol for Pigs. Welfare Quality Consortium, Lelystad, The Netherlands.
- Welfare Quality®, 2009c. Welfare Quality® Assessment Protocol for Cattle. Welfare Quality Consortium, Lelystad, The Netherlands.
- Wemelsfelder, F., 1997. The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science* 53, 75–88.
- Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T., Lawrence, A.B., 8/2001a. Assessing the “whole animal”: a free choice profiling approach. *Animal Behaviour* 62, 209–220.
- Wemelsfelder, F., Lawrence, A.B., 2001b. Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. *Acta Agriculturae Scandinavica - Section A: Animal Science* 21–25.
- Wemelsfelder, F., Millard, F., De Rosa, G., Napolitano, F., 2009c. Qualitative Behaviour Assessment, Assessment of Animal Welfare Measures for Dairy Cattle, Beef Bulls and Veal Calves. Welfare Quality, Lelystad.
- Wemelsfelder, F., Nevison, I., Lawrence, A.B., 2009d. The effect of perceived environmental background on qualitative assessments of pig behaviour. *Animal Behaviour* 78, 477–484.

Whay, H.R., Main, D.C.J., Green, L.E., Webster, A.J.F., 2003. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. *Animal Welfare* 12, 205–217.

Winckler, C., 2009. Verhalten der Rinder. In: Hoy S. (Hrsg.), *Nutztierethologie*. Ulmer, Stuttgart.

## Abbildungsverzeichnis

Abbildung 1: Verteilung der Begriffe im Komponentendiagramm.....	31
Abbildung 2: Verteilung der Begriffe im Komponentendiagramm. Komponentenladung gibt die Korrelation der Begriffe mit den Dimensionen 1 und 2 an. Betriebe beurteilt durch 5 BeobachterInnen.....	36

## Tabellenverzeichnis

Tabelle 1: Die „Welfare Quality® Prinzipien“ und „Welfare Quality® Kriterien“ als Basis des Welfare Quality®-Erhebungsprotokolls (Welfare Quality®, 2009c).....	10
Tabelle 2: Betriebsdaten.....	17
Tabelle 3: Kendall's Konkordanzkoeffizient W für 13 BeobachterInnen für die Parameter PCA QBA-Scores für Dimension 1 und Dimension 2 sowie WQ QBA- Score zu verschiedenen Tageszeiten.....	27
Tabelle 4: Kendall's Konkordanzkoeffizient W der QBA Parameter PCA QBA- Scores von Dimension 1 und Dimension 2 sowie WQ QBA-Score zu verschiedenen Tageszeiten aus der Beurteilung von 5 BeobachterInnen.....	28
Tabelle 5: Mittelwert ( $\pm$ SD) WQ QBA-Score für 10 Betriebe beurteilt von 13 BeobachterInnen.....	29
Tabelle 6: Mittelwert ( $\pm$ SD) PCA QBA-Score von Dimension 1 für 10 Betrieben beurteilt von 13 BeobachterInnen.....	32
Tabelle 7: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 2 für 10 Betrieben beurteilt von 13 BeobachterInnen.....	33
Tabelle 8: Mittelwert ( $\pm$ SD) WQ QBA-Score für 10 Betriebe beurteilt von 5 BeobachterInnen.....	34
Tabelle 9: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 1 für 10 Betriebe beurteilt von 5 BeobachterInnen.....	37
Tabelle 10: Mittelwert ( $\pm$ SD) PCA QBA-Score für Dimension 2 für 10 Betrieben beurteilt von 5 BeobachterInnen.....	38

# 10 Anhang

Erhebungsblatt für die Videobeurteilungen:

## Bewertungsskala für die qualitative Verhaltensbeurteilung von Milchrindern (QBA)

**Observer:**

**Date:**

**Farm:**

**Tageszeit:**

**Haltungssystem:**

**Anzahl der Tiere:**

**Rasse:**

**Kurze Beschreibung des Haltungssystems bzw. der Stalleinheit (z.B. Innen-/Außenbereich, Einstreu, Stalleinrichtungen, Licht, Fütterung, etc.):**

---

Bitte beobachten Sie die Tiere im Haltungssystem für 10 bis 20 Minuten und beurteilen Sie dann den Ausdruck des gesehenen Verhaltens ('Körpersprache'):

aktiv	Min.	_____	Max.
entspannt	Min.	_____	Max.
ängstlich	Min.	_____	Max.
aufgewühlt	Min.	_____	Max.
ruhig	Min.	_____	Max.
zufrieden	Min.	_____	Max.
gleichgültig	Min.	_____	Max.

frustriert	Min.	_____	Max.
freundlich	Min.	_____	Max.
gelangweilt	Min.	_____	Max.
verspielt	Min.	_____	Max.
positiv beschäftigt	Min.	_____	Max.
lebhaft	Min.	_____	Max.
neugierig	Min.	_____	Max.
leicht irritierbar	Min.	_____	Max.
‘unentspannt’	Min.	_____	Max.
gesellig/ ‘sozial’	Min.	_____	Max.
teilnahmslos	Min.	_____	Max.
glücklich	Min.	_____	Max.
gestresst	Min.	_____	Max.

Allgemeine Kommentare oder Beobachtungen:

## **11 Ehrenerklärung**

Ich erkläre, dass ich die Arbeit selbst verfasst, alle Quellen zitiert und mich keiner unerlaubten Hilfsmittel bedient habe.

17. Januar 2014