University of Natural Resources and Life Sciences, Vienna

Department of Sustainable Agricultural Systems

Division of Livestock Science

Animal Husbandry Group

Comparison of Free Choice Profiling and Fixed Terms for Qualitative Behaviour Assessment In Dairy Cattle

MASTER THESIS

by

Bakk.rer.nat. Beate Müllner

Wien, 2011

Advisors: Univ.Prof. Dr.med.vet. Christoph Winckler
Dipl.Biol. Anke Gutmann
Dr. Christine Leeb

CONTENT

1. INTRODUCTION 3

1.1. AIMS AND HYPOTHESIS 4

1.2. MEASURING ANIMAL BEHAVIOUR 5
1.2.1. Quantitative Measures of Behaviour 5
1.2.2. Qualitative Measures of Behaviour 6
1.2.2.1. Free Choice Profiling (FCP) 7
1.2.2.2. Fixed Term Assessing (FT) 11

2. MATERIAL AND METHODS 13

2.1. VIDEO RECORDINGS 13

2.2. STUDY DESIGN 13
2.2.1. Free Choice Profiling 14
2.2.2. Fixed Terms 15

2.3. STATISTICAL ANALYSIS 17

3. RESULTS 19

3.1. COMPARISON OF FCP AND FT USING INDIVIDUAL ANIMAL CLIPS 19

3.2. COMPARISON OF FCP AND FT USING HERD CLIPS 24

3.3. COMPARISON OF INDIVIDUAL ANIMAL AND HERD ASSESSMENT 28

4. DISCUSSION 30

5. CONCLUSION 34

6. SUMMARY 35

7. ZUSAMMENFASSUNG 36

8. APPENDIX 37

9. REFERENCES 59

1. Introduction

Animal welfare assessment has become of increasing importance in livestock production because customers increasingly appreciate high levels of welfare standards for farm animals they buy products from. Animal welfare can be defined as a "state of harmony between an individual and its environment" (Désiré et al., 2002). Figure 1 illustrates the current concepts of animal welfare (e.g. Lund, 2002). It demonstrates that animal welfare consists of three main elements - physical health and subjective state as well as the ability to perform natural behaviour (Lund, 2002). The FAWC (1992) provided five basic requirements to achieve animal welfare: freedom from hunger and thirst, from discomfort, from pain, injury or disease; freedom to express normal behaviour and freedom from fear and distress.

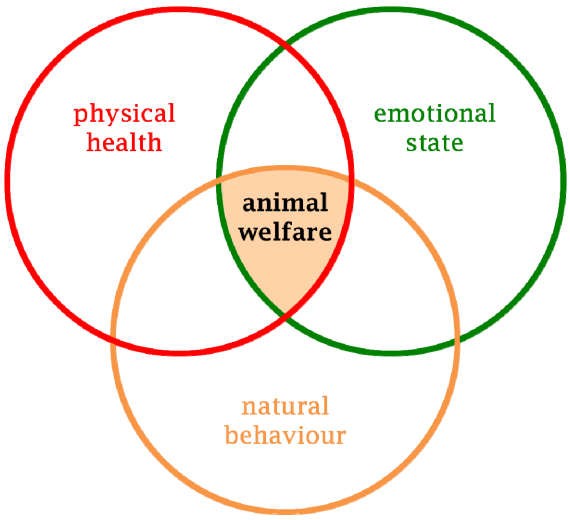


Figure 1: Definition concept of animal welfare (Lund, 2002).

One of the most well known definitions described animal welfare as "its state as regards its attempts to cope with its environment" (Broom, 1986). Measuring animal welfare relies on a variety of parameters and methods with a focus on animal-based measures. Behavioural measures therefore play an important role (Broom, 1991; Rousing and Wemelsfelder, 2006). However, Rousing and Wemelsfelder (2006) state that quantitative behaviour evaluation may omit potentially important information. Qualitative Behaviour Assessment might therefore be a new approach to complement quantitative behavioural assessment methods. It is usually conducted using the Free Choice Profiling (FCP) approach. More recently, a to some extent simplified

version using so called Fixed Terms (FT) has been proposed, especially for on-farm assessments. However, to our knowledge, to date the agreement of both approaches in terms of inter-observer reliability and qualitative content of the results has not been investigated.

1.1. Aims and Hypotheses

The overall aim of this study was to compare two methods of generating terms for Qualitative Behaviour Assessment in dairy cattle: Free Choice Profiling and Fixed Terms. Whilst Free Choice Profiling is mainly used to evaluate groups of animals or herds (Wemelsfelder et al., 2009c), Qualitative Behaviour Assessment has often been applied to individual animals (Napolitano et al. 2008, Kuhar et al. 2006). Therefore, the second aim was to investigate if Free Choice Profiling can be used for assessing groups of animals as well as individual animals.

The following hypotheses have been formulated:

a) Free Choice Profiling and Fixed Terms achieve a similar agreement between observers

b) Independently from the method, assessing individual animals achieves a higher inter-observer agreement than assessment of groups of animals

c) Free Choice Profiling and Fixed Terms reveal comparable dimensions to describe animal behaviour in both individual animals as well as groups of animals

d) Free Choice Profiling and Fixed Terms reveal a good correlation of the dimension scores

1.2. Measuring Animal Behaviour

The scientific evaluation of animal behaviour is an important tool to obtain information about the species; it allows finding out an animal's behaviour and can be used to analyse scientific questions and hypotheses (e.g. Broom, 1997; Dawkins, 2003). Only if we know as much as possible about animal behaviour it will be possible to improve animal husbandry and animal welfare. Basically, there are two main measurement methods: Quantitative and Qualitative Behaviour Assessment.

1.2.1. Quantitative Measures of Behaviour

Using quantitative assessment methods, for example incidence and duration of behaviours are recorded. It is therefore possible to describe how often or how long an animal shows a specified behaviour (Nelson et al., 2009), but details of animal's wellbeing cannot revealed. Cook et al. (2007) demonstrated the limited possibilities of quantitative assessment methods. The scientists described behaviour changes in different climatic conditions. Fourteen cows in four filming sessions were observed under different climatic conditions to find changes in the time budget of lactating dairy cows. The mean of the lying time decreased from nearly eleven to nearly eight hours per day from the coolest to the hottest session filmed (Cook et al., 2007). Drinking and standing in an alley increased with increasing temperature. This study demonstrates a typical quantitative behaviour assessment. The scientists noticed changes in behaviour over time affected by changes in temperature. Cook et al. (2007) only demonstrated changes in behaviour, but did not give information about the animals' emotional state or their wellbeing under different climatic conditions; the scientists did not show up "how" the animals change their behaviour – there was no statement about the welfare of these dairy cows under different climatic conditions. One way to overcome this lack of information is the use of Qualitative Behaviour Assessment, which will be described in the following sections.

1.2.2. Qualitative Measures of Behaviour

Wemelsfelder (1997) mentioned that an assessment methodology is needed to evaluate behaviour with terms of an "individual's perspective". Qualitative Behaviour Assessment is this "subject-based approach" to evaluate animal's behaviour "as a whole", because the "animal as a whole is the dynamic, integrative centre of action" (Wemelsfelder, 1997). Qualitative Behaviour Assessment bases on integrating measurement and interpretation (Rousing and Wemelsfelder, 2006). The qualitative approach makes it possible to find out subtle details about movement, posture and changes in behaviour over time and aspects of the context in which the behaviours occur (Wemelsfelder et. al, 1997; 2000). This information about the body language shows "how" an animal evaluates its present situation (Wemelsfelder, 1997).

Rousing and Wemelsfelder (2006) described Qualitative Behaviour Assessment as a method based upon the integration by observers of perceived behaviour expression. These qualitative evaluations describe animal behaviour with terms such as "nervous", "calm" or "aggressive" (Wemelsfelder et al., 2000; Rousing and Wemelsfelder, 2006). Thus, Qualitative Behaviour Assessment does "not evaluate at all what an animal does, but how it does what it does" (Wemelsfelder et al., 2000). The Qualitative Behaviour Assessment summarises the different aspects of the animal's "dynamic style of interaction" with its environment (Wemelsfelder et al., 1997; 2000; 2001a, 2001b).

By now, two different methods have been developed to generate terms for Qualitative Behaviour Assessment: Free Choice Profiling, where the panellists generate their own terms to assess the animals and the so called Fixed Terms approach, which uses predefined descriptors.

1.2.2.1. Free Choice Profiling (FCP)

Free Choice Profiling (FCP) is a common method of subjective evaluations and an approach to profile analysis in which each observer produces individual terms to describe a sample or an animal. Basic principles of Free Choice Profiling are that several observers generate their own adjective terms that are subsequently used to evaluate samples. Observers are completely free in choosing their descriptive terms. It is a long-known assessment method first used in sensory studies.

Williams and Langron (1984) demonstrated the technique of Free Choice Profiling for the olfactory assessment of commercial port wine. The authors described Free Choice Profiling as a “natural way” to assess products, because people can use their own terms. The panellists had to list sensory characteristics, which describe the sensory properties of the ports. The assessors developed between six and eighteen words. The study by Williams and Langron (1984) demonstrated that it is not necessary to use precisely defined words for describing samples to show relationships and differences. Although, panellists used different words for describing the wines, they showed a high similarity in their evaluation, because there was approximately the same level of agreement or disagreement in flavour and aroma (Williams and Langron, 1984).

Sinesio and Moneta (1997) described sensory assessment as an increasingly used tool to evaluate foodstuff. For example, Thamke et al. (2009) used the Free Choice Profiling approach to evaluate six formulations of dark chocolate. 39 panellists with different local backgrounds had to judge the sensory characteristics of the chocolate samples using their own vocabulary. The assessors had to evaluate taste, flavour and mouth feeling. They generated between three and thirteen terms for sensory characterization. The scientists suggested that the test persons with different backgrounds used an identical vocabulary; they found the same “key descriptors” to describe the samples (Thamke et al., 2009). The panellists showed a good agreement by arranging the samples based on the sensory data. Panellists evaluated the mouth feeling of chocolate with low cocoa content as melting and creamy, chocolate with high cocoa content as mealy, sticky and dry (Thamke et al., 2009). Such subjective details could not be revealed with quantitative assessment methods. The scientists also mentioned that panellists seemed to be limited in generating different terms, because the test persons with different local backgrounds used the same “key de-

scriptors". This indicates the high similarity of subjective evaluations by different observers and that sensory evaluations are amenable for scientific investigations.

The facts that qualitative assessments bring out hidden details and that different assessors from different countries use identical words to describe these details make qualitative assessments an interesting method for gathering details about animal welfare. In 1997, Wemelsfelder mentioned that animals are treated as subjects in interaction with humans and described this fact with an example: People use the anthropomorphic language and say - probably the cat is scratching on the door - "the cat wants to get out" Therefore, the subjective assessment of animal's behaviour is called as a direct conception (Wemelsfelder, 1997) and animal behaviour is amenable for qualitative evaluations. Qualitative Behaviour Assessment might be able to detect hidden information about the animal behaviour that could not be found out with quantitative methods, it is also used to find out details about animals welfare. Especially the Free Choice Profiling approach enables an independent evaluation of animal behaviour.

Rousing and Wemelsfelder (2006) stated that Qualitative Behaviour Assessments seem to be called in question, because of the commonly assumed uncertain validity. Many scientists may cast doubt on Qualitative Behaviour Assessment, because that may have an anthropomorphic view on animal behaviour, but Wemelsfelder adheres in several papers (2000; 2001a; 2001b; 2006) that qualitative evaluations of behaviour are amenable to scientific analysis and that it is based on observable parameters.

An often-discussed question is the differences between lay people and experts. Some studies investigated the effect of level of expertise (e.g. regarding animal production and animal behaviour) on the results obtained from Qualitative Behaviour Assessment. Foley (1935) investigated the "judgment of facial expression of emotion in the chimpanzee". 127 students had to assess photographs of chimpanzee with sixteen given terms. One picture showed a grinning chimpanzee and the students interpreted this as a sign of joy and laughter; in reality the behaviour of the monkey showed defensive threat or appeasement. For that reason, several studies investigate the results of Qualitative Behaviour Assessments with trained or untrained people. Meagher et al. (2009) investigated if trained panellists may minimize the error-proneness of Free Choice Profiling because untrained persons may tend to use words that describe a human state. They write that humanized terms may not be ap-

propriate for describing a similar state in an animal. This fact is described as the risk of allowing the untrained observers to generate their own terms (Meagher, 2009). Meagher (2009) also noted that experience does not improve rating validity. It would be favourable if the panellists were well trained in the Free Choice Profiling procedure and if they had some experience in that specific area (e. g. pain scoring; Meagher, 2009).

Wemelsfelder et al. (2000) mentioned that it is always possible that if observers agree they can be wrong. They explained that different terms might have close meanings (e. g. “fending”/”agitated”). But they also indicated that terms similar in tone might reflect different aspects of the expressive behaviour. To minimize this error-proneness it seems to be recommendable to use panellists with a very good knowledge of the species that is chosen for assessment. Wemelsfelder et al. (2000) demonstrated the spontaneous qualitative assessment of behavioural expressions in pigs with untrained panellists. Eighteen naïve assessors had the opportunity to qualitatively characterize, independently and with their own words, the behavioural expressions of twenty individual growing pigs (Wemelsfelder et al., 2000). The pigs were singly in a pen and could interact with a human in the centre of this pen. After observing the pigs first, the observer had to write down their terms that summed up the qualities of the animal’s behaviour (Wemelsfelder et al., 2000). This procedure was repeated one month later with twenty different pigs. The observers were not experienced in observation of pigs and not trained in describing behaviour but they achieved a significant agreement when given freedom to describe the pig’s behaviour with their own terms (Wemelsfelder et al., 2000).

Observer evaluations are only valuable if the observers generate reliable and valid data (Meagher, 2009). Therefore, several studies investigated the inter- and intraobserver reliability. Inter-observer reliability is defined as “agreement between multiple people independently rating the same individual” (Meagher, 2009). Intra-observer reliability is defined as “agreement between ratings by the same individual on multiple occasions” (Meagher, 2009). In 2006, Rousing and Wemelsfelder evaluated the inter- and intraobserver reliability and its correlation to quantitative ethogram-based assessment. Twelve experienced observers had to assess 25 video clips showing various types of cow behaviour at the drinker by using Free Choice Profiling (Rousing and Wemelsfelder, 2006). Additionally, an ethogram that included frequency and duration of several behaviour expressions was used to quantify the

cows' behaviour. The observers showed a significant agreement in their assessment. They qualified video clips as "aggressive"/"bullying", which showed frequent head butting. The frequently shown social licking interactions were qualified as "relaxed"/"calm" (Rousing and Wemelsfelder, 2006). For this reason, the scientists revealed that Qualitative Behaviour Assessment is a reliable method for on farm assessment of social interactions of dairy cattle.

Quantitative Behaviour Assessment evaluates the incidence and duration of animal behaviour while Qualitative Behaviour Assessment evaluates the "animal as a whole" (e.g.: body language) (Rousing and Wemelsfelder, 2006). Qualitative Behaviour Assessment seems to be the assessment method that supports the quantitative methods best. For this reason, it seems advisable to conduct both methods in parallel, which was done in several studies. Rousing and Wemelsfelder (2006) mentioned that the combination of both methods might help to identify key-indicators that assist in the assessment of animal behaviour.

Minero et al. (2009) combined qualitative and quantitative measurements to assess the response of foals to an unfamiliar human. The quantitative evaluation was done using instantaneous sampling; the qualitative analysis was conducted with Free Choice Profiling. To compare the qualitative and quantitative assessment data Principal Component Analysis and the Spearman's rank correlation were used. Minero et al. (2009) demonstrated a highly significant score between the correlation of the quantitative and qualitative Factor 1 that indicated that the foals engaging in close contact with the experimenter were assessed as "explorative"/"sociable". Horses that showed immobility behaviour were described as "suspicious"/"nervous". The scientists found a link between quantitative and qualitative assessment by evaluating foals behaviour associated with an unfamiliar human. This study confirms a useful combination of quantitative and qualitative assessments (Minero et al., 2009). Thus, the scientists suggested that the traditionally retrospective, quantified assessment might be supplemented by a qualitative approach. Many studies (Rousing and Wemelsfelder, 2006; Minero et al., 2009) illustrated that it is possible and useful to combine the qualitative behaviour methods with quantitative assessments.

1.2.2.2. Fixed Term Assessing (FT)

Free Choice Profiling seems to be unsuitable for on-farm assessment (Wemelsfelder et al., 2009d). It is a complex and time-consuming procedure. For a successful assessment, several persons are needed to generate own terms and the terms have to be generated in advance in a time consuming procedure. The fact that Free Choice Profiling is not very feasible for on-farm assessment stimulated the development of a standard assessment protocol proposing Fixed Terms lists. The Fixed Term assessment method is conducted with given descriptors and simplifies on-farm assessment. Fixed Terms were already used to investigate personality and behaviour expressions of different species. They simplify behaviour assessments and allow quick conduction because no complex “term-finding-phase” is needed.

Fixed Terms are used in different areas of animal behaviour assessments. In 1999, Wielebnowski studied behavioural differences as predictors of breeding status in captive cheetahs using the Fixed Terms approach. 25 female and 19 male cheetahs were evaluated using eighteen qualitative terms – active, aggressive to conspecifics, aggressive to people, calm, curious, eccentric, excitable, friendly to conspecifics, friendly to people, fearful of conspecifics, fearful to people, insecure, playful, self-assured, smart, solitary, tense and vocal. The chosen terms to assess the cheetahs were based on studies with other species (domestic cats and rhesus monkeys). For that reason, the terms were modified (Wielebnowski, 1999). The data were analysed with Principal Component Analysis. Three major components were found that divided the animals in tense-fearful, vocal-excitable, and aggressive (Wielebnowski, 1999). Additionally to a questionnaire the direct behaviour observations might be a useful approach to investigate breeding problems and „behavioural idiosyncrasies in captive-held species“ (Wielebnowski, 1999).

In 2004, the EU project Welfare Quality® started, which had – among others – the aim to develop an on-farm welfare assessment standard and practical measures to improve animal welfare (Wemelsfelder et al., 2009c). Within Welfare Quality® assessment protocols for three livestocks were developed: poultry (Wemelsfelder et al., 2009a), pigs (Wemelsfelder and Millard, 2009b) and cattle (Wemelsfelder et al., 2009c). For the development of Fixed Terms as qualitative descriptors of dairy cattle behaviour 22 groups of dairy cattle were evaluated with four assessors. The first tested list contained 26 terms. After feedback from research partners in the welfare

quality cattle subgroup, the adjusted list contained 29 terms. Because the study was conducted in Italy, the original English terms had to be translated to Italian. As done in several trials (Wemelsfelder et al., 2000; Thamke et al., 2009; Villarino, 2004), the assessors got a detailed introduction on how to carry out Qualitative Behaviour Assessment of a herd or a group of animals (Wemelsfelder et al., 2009c). The number of descriptors was finally reduced to twenty. So the list might be more practicable for on-farm assessment (Wemelsfelder et al., 2009c). The removing based on three principles (Wemelsfelder et al., 2009c):

1) removed terms had a "low loading on each of the assessors' Principal Components Analysis components";
2) removed terms had "no meanings with clear relevance to welfare";
3) removed terms had "similar meanings with other used terms".

Anyway, Wemelsfelder et al. (2001a; 2001b) favour Free Choice Profiling because provided terms may impair the independence of the observers, so that the ratings would not reflect their independent evaluations of the behaviour. The observers are forced to project preconceived descriptors to evaluate the animal behaviour. For that reason, the integrative character of qualitative assessment is prejudiced and obstructed (Wemelsfelder et al., 2000). Meagher (2009) agreed with Wemelsfelder that provided terms would restrict the descriptors' selection.

Another risk may be the chosen terms for Fixed Term assessment. Meagher (2009) stated that the choice of terms should involve careful thought, taking into account the knowledge about the species. This indicates that it is necessary to create different standard protocols for different species to describe their specific behaviour. This was realised during the development of these assessment protocols by experts that is explained detailed in the Welfare Quality Reports (e.g.: Wemelsfelder et al., 2009c).

2. Material and Methods

There were two rating sessions, which took place in April 2010. It was necessary to prepare forty video clips and the respective assessment protocols.

2.1. Video recordings

Twenty individual animal clips (i-clips) and twenty herd clips (h-clips) were used. The h-clips showed different breeds of cows (Holstein Friesian, Brown Swiss, Simmental Breeds) in loose housing systems. Each clip lasted for one minute and was shown without sound. The clip content was chosen to cover a diversified behaviour spectrum. The h-clips were in colour with superior image quality.

The i-clips showed Simmental cows in a deep litter system during social licking interactions. To ensure that only the receiver of this social interaction was recognizable and subject to scientific assessment, the other cows were masked with a black screen using the Adobe Premiere software. All i-clips lasted for one minute and were black and white. The quality was moderate and the clips were shown without sound. The clips had been selected as to represent a high diversity of dairy cattle reactions to social licking.

2.2. Study Design

Four male and eight female observers were available for the Qualitative Behaviour Assessment sessions. All observers were familiar with the basics of dairy cattle behaviour, but there were only four people with experience in Qualitative Behaviour Assessment. As in studies carried out by Wemelsfelder et al. (2000, 2001a. 2001b), all observers received a detailed introduction, which provided information about the procedures they had to follow. In total there were two rating sessions, which were one week apart in order to minimise carry-over effects from the previous experiences.

2.2.1. Free Choice Profiling

On the first day, the observers were introduced to Free Choice Profiling. Free Choice Profiling was chosen as the first method they were introduced to in order to ensure that panellists were not influenced by predefined terms. First, introduction to Free Choice Profiling provided information about the general approach of Qualitative Behaviour Assessments and aims of the study. Using pictures, the panellists then received a brief training in generating terms for expressive styles of behaviour. This was followed by an explanation of the principles of assessment of groups of animals such as avoiding to focus on individual cows.

After this general introduction all observers generated their own terms by watching ten clips of each kind of clips. Presentation of clips started with ten i-clips followed by ten h-clips. Each clip was shown separately. After each clip, the observers separately noted down the terms of expressive behaviour quality they associated with the situation shown. For this purpose, a form had been designed to simplify the collection of terms for each clip. There was no time limit to analyse the clip content and write down terms. As described by Wemelsfelder et al. (2000, 2001a, 2001b), the observers were requested not to discuss the terms. This ensured the independence of each observer.

In a second step, the individual terms were entered in a Free Choice Profiling list by observer. In the case of terms with a very close meaning one term was excluded (“enjoying”/”appreciative”) (“genießend”/”genüsslich”); this was also the case if terms described both ends of a continuum such as “active”/”inactive” (“aktiv”/”inaktiv”). Furthermore terms without an expressive quality were also removed (e.g. “running” [“laufen”]). In the end, for each observer an individual scoring protocol comprising all terms that had been created was developed. Every single term on the assessment protocols had a 12.5 cm long Visual Analogue Scale (VAS), which ranged from minimum (animals do not show the attribute) to maximum (the highest level of the attribute). The assessment protocols and generated Terms of each observer are listed in appendix. The Free Choice Profiling lists contained a different number of terms, which probably depended on the creativity of the observer.

In the afternoon of the same day, the rating started with Free Choice Profiling by using the Free Choice Profiling protocols created in advance. The twelve observers were split into two groups of six persons each. Both groups consisted of four

women and two men to ensure gender distribution. The two groups watched the clips in different rooms. The first group started to score the i-clips while the second group scored the h-clips first. After watching each one-minute-clip the observers ticked every scale at a point between minimum and maximum. Every single term had to be assessed. It was very important that the observers do not miss out even a single term. Basically, there was no time limit to tick the scales. This was necessary since the creative assessors had to score more terms and therefore they needed more time.

2.2.2. Fixed Terms

One week later, the second session took place using Fixed Terms for the assessment. The Fixed Term list for the herd assessment was taken from the Welfare Quality assessment protocol (Wemelsfelder et al., 2009c). The twenty terms were translated into German. Furthermore, a Fixed Term list for the individual animal rating had to be created because such an assessment protocol to observe an individual animal in a specific situation did not exist before. Since the i-clips showed a social licking interaction and assessment should focus on the receiver of this interaction, the Welfare Quality list served only as the basis. The terms “positively occupied” (“positiv beschäftigt”), “curious” (“neugierig”) and “sociable” (“gesellig”) were replaced by the terms enjoying” (“genießend”), “requesting” (“auffordernd”), and “intrusive” (“aufdringlich”). All the other terms remained unchanged. As done with the Fixed Terms for herd assessment, the twenty terms on the individual animal list were translated into German.

The introduction to the Fixed Terms rating took place with all twelve observers before the rating started. To ensure that all twelve observers understood the same meaning of every single term, all terms were discussed. E.g. the term “curious” (“neugierig”): Body language and behaviour of curious cows were discussed. The principles of herd assessment were repeated. As a result of this detailed introduction, the observers were assumed to be able to assess the herds without focusing on any individual cow. The clips were performed in different sequence in both ratings. As for Free Choice Profiling, every Fixed Term had a 12.5 cm long scale ranging from minimum to maximum, which should be marked after watching the video clips. As done in the Free Choice Profiling there was no time limit to tick the scales.

Table 1 contains information on both rating sessions. Important differences are emphasised and summarised. First, the introduction to Free Choice Profiling and to Fixed Terms provided different information, respectively. The procedure to generate terms is not required for Fixed Terms. For that reason, the development of the Free Choice Profiling scoring protocols deviated from the development of the Fixed Term scoring protocols. The group formation was the same on both session days. The Fixed Term scoring happened one week later with a different clip sequence and the groups started with different types of clips (i-clips, h-clips) on both rating days.

Table 1: Summary of the study design.

1. Session	**Free Choice Profiling**
FCP-Introduction	Detailed introduction to all observers about Free Choice Profiling and the procedure; herd assessment principles: give an overall assessment without focussing on individual cows
Generating terms	All observers create their own terms by watching ten clips of each type of clips (i.e. herd and individual animal)
Free Choice Profiling protocols	Exclusion of terms describing equal qualities; terms were entered in the personal assessment protocols of each observer
Observer groups	Splitting the 12 observers into two groups of 6 persons each – 4 women and 2 men
Scoring	Scoring the clips using the FCP-protocols that had previously been generated. One group started with the i-clips, the other group with the h-clips
2. Session	**Fixed Terms** **One week later; clip sequence of both types of clips was modified**
Fixed Terms - Introduction	Clarification of the Fixed Terms meanings, repeat of the herd assessment principles: do not focus on any cow – give an overall assessment
Fixed Terms protocols	Herd assessment with the Welfare Quality Protocol; assessment of the individual animal clips with the modified WQP
Observer groups	Groups consisted of the same people as in session one
Scoring	Scoring the clips with the Fixed Terms. Groups which had started with h-clips in the first rating session started with i-clips and vice-versa

2.3. Statistical Analysis

The Fixed Term data as well as the Free Choice Profiling data were analysed using Generalised Procrustes Analysis (GPA) in order to have the same statistical basis. The first step was to measure the distance in millimetres on the rating scales of each term from minimum to the placed mark. The same ruler was used to avoid measurement errors. If a mark had been placed between e.g. 38 and 39 millimetres, a value of 38 millimetres was used. The next step was to create data matrices for each assessor in MS Excel. Following this preparatory work, the actual statistical analysis was done. The Generalized Procrustes Analysis feature of the software package GenStat 14 was used.

General Procrustes Analysis is described as a multivariate statistical technique (Gower, 1975) and it is used to analyse sensory Free Choice Profiling data (Gower, 1975; De Jong et al., 2002; Wilkinson et al., 2000). Wemelsfelder et al. (2001a) described General Procrustes Analysis as a ”pattern-matching mechanism”. Although the observers use different terms, it is possible to compare the samples (in that case: video clips), because these samples are constant (Xiong et al. 2008). General Procrustes Analysis transforms each data matrix into “multidimensional configurations”, which makes it possible to compare the matrices of the assessors (Xiong et al., 2008); in complex geometric transformation steps General Procrustes Analysis determines the similarity between these configurations (Wemelsfelder et al., 2000).

General Procrustes Analysis uses “an iterative algorithm to find rotation and transformation matrices and scaling factors which minimize some measure of the distance between each matrices, the loss function” (Wilkinson et al., 2000). Several transformation steps are necessary to find a “best common denominator”, which is called the “consensus profile” (Wemelsfelder et al., 2000). An add-on randomisation test (Wu et al., 2003) provides information about this consensus profile as a significant feature of the data set or an artefact of the process of transformations (Wemelsfelder et al., 2001a). General Procrustes Analysis allows detecting the level of consensus between all assessment patterns (Wemelsfelder et al., 2001a). Therefore, the “basis of the multidimensional intersample distances specified by each observer” is used (Wemelsfelder et al., 2001a).

Several plots are the outcome of General Procrustes Analysis. The so-called Agreement Plots demonstrate the Procrustes Statistic, which Wemelsfelder et al.

(2000; 2001a;) called “goodness-of-fit” between the observer matrices (Wemelsfelder et al., 2000; 2001a). A high value of the Procrustes Statistic indicates a good observer agreement (Wemelsfelder et al., 2000). These Agreement Plots demonstrate the relative distance between the observers and the consensus profile (Wemelsfelder et al., 2000).

The so-called Word Charts are also outcomes of the General Procrustes Analysis. As explained in studies done by Wemelsfelder et al. (2000, 2001a, 2001b), every single graphic Word Chart created by General Procrustes Analysis shows the terms generated by the assessors as descriptors of the main consensus dimensions. All terms of each assessor are plotted on the x- and y-axis. As described in Wemelsfelder et al. (2000), “the axes reflect the scaling values for relative observer distance”. There are as many dimensions as descriptors (Wemelsfelder et al., 2001a). The dimensions are shown on the Word Chart plots as axes that reflect the main dimensions of the consensus profile and indicate which of these terms “best correlate with those axes” (Wemelsfelder et al., 2000). A high correlation gives a good characterization of the dimension (Wemelsfelder et al., 2000).

To find out the main descriptors over all assessors, all Word Charts were analysed. All terms created by an assessor can be found on the Word Chart of this assessor. The terms have a value, which gives the correlation on the dimensions. It was decided that the dimensions were characterised with terms having a high loading value more than 0.5. If there were several terms with a high or higher loading of 0.5 the frequency of mentioning was used as restriction. Normally, it is possible to explain more than 60% of the variance with the first two dimensions. Therefore, the first two dimensions were used as the main dimensions. But in some cases, these main dimensions did not explain more than 60%, so it was sometimes necessary to include the third dimension.

Concerning the actual assessment of the clips, General Procrustes Analysis produces consensus scores on the main dimensions for every single clip and a corresponding Score Plot. This plot illustrates the main dimensions – as explained above – and the position of each clip on these dimensions. A circle on these Scores Plots illustrates a standard error for each clip position on the plot. Encompassing only one clip at a time indicates that the position of this clip on the plot is reliably fixed by its score on each of the axes of the consensus profile (Wemelsfelder et al., 2000).

To demonstrate a correlation between the dimensions, a Pearson correlation test was used to compare scores. The Person Correlation gives a value between -1 and +1. This is a measure for linear correlation between variables.

3. Results

First, the comparison of assessment methods - Free Choice Profiling and Fixed Terms - is demonstrated. GPA plots are presented for better illustration. Subsequently, the results of the individual and herd assessments are presented.

3.1. Comparison FCP and FT using Individual Animal Clips

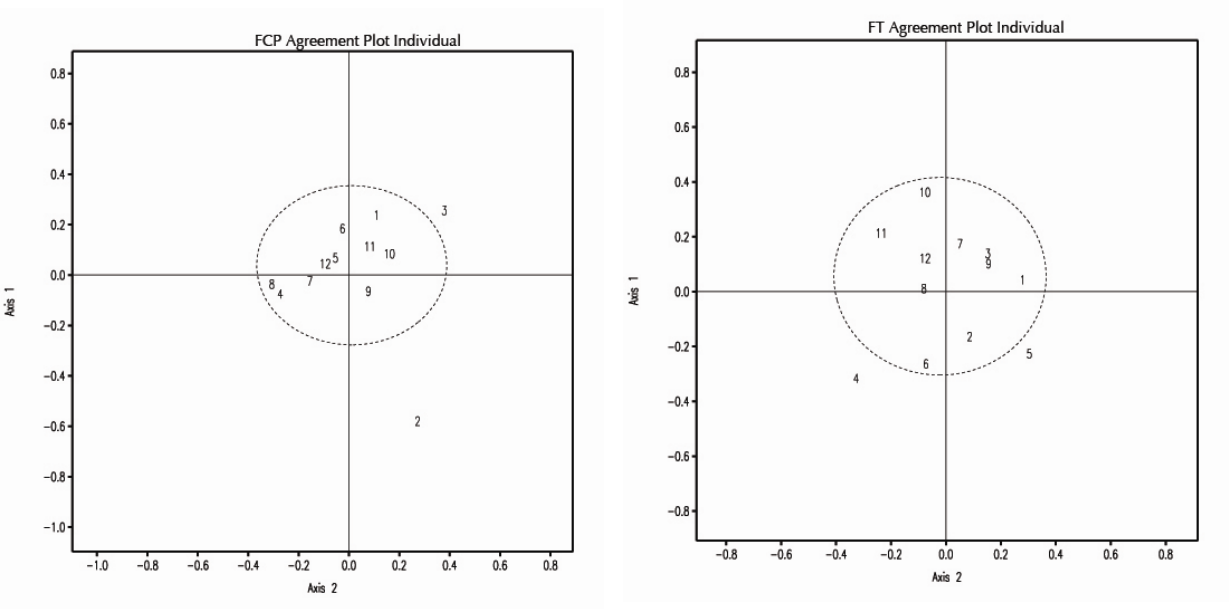


Figure 2: FCP and FT observer agreement for the i-clips (n=12 observers)

The Agreement Plots (Figure 2) illustrate the Procrustes Statistic of the twelve assessors. With both Qualitative Behaviour Assessment methodologies two outliers were located outside of the 95%-confidence region. The Free Choice Profiling plot of the i-clip assessment showed one observer located very far from the confidence region. This indicated a lower agreement with the other assessors. The other outlier on the Free Choice Profiling plot was close to the confidence region. The Procrustes Statistic of Free Choice Profiling was 57.96%. The mean of 100 randomised profiles was 42.48% of the variation between matrices. The simulated variance had a value

of 0.3909. The one-sided t-test showed a significant difference between the total percent variation explained by the consensus and the simulated mean (df = 99, t = 24.8, p<0.001; Table 2).

On the Fixed Terms Agreement Plot the outliers are located close to the confidence region. The Procrustes Statistic of Fixed Terms was 56.93% of the total variation; the mean of the 100-randomised profiles had a value of 42.94% of total variation between the matrices. The simulated variance had a value of 0.5645; the one-sided t-test also showed a significant difference between the total percent variation explained by the consensus and the simulated mean (df = 99, t =18.6, p<0.001; Table 2).

Table 2: Comparison of FCP and FT Procrustes Statistics for i-clips

Individual Animal Assessment		
GPA results	**FCP**	**FT**
Consensus profile	57.96%	56.93%
Randomised profile (mean ± variance)	42.48% ± 0.39	42.94% ± 0.56
t_{99} significance test	24.8 (p<0.001)	18.6 (p<0.001)

The Word Charts demonstrated the characterization of the main Dimensions 1 and 2 (see appendix for all Word Charts). As explained above, the most frequently used terms with a loading of at least 0.5 were chosen. The most frequently generated terms of each assessor are listed in Table 3. These terms were entered in the Score Plots mentioned below (Figure 3). During Free Choice Profiling, several assessors generated identical terms as provided in the Fixed Terms Welfare Quality protocol.

Table 3: Most frequently used terms with a loading of ≥ 0.5 in FCP and FT assessment of i-clips (number of times the term was used in brackets):

Correlation of dimension	FCP	FT
d1 positive correlation	indifferent (gleichgültig) (6)	bored (gelangweilt) (6)
d1 negative correlation	enjoying (genießend) (3), **requesting** (auffordernd) (2)	**requesting** (auffordernd) (4)
d2 positive correlation	**relaxed** (entspannt) (2), calm (ruhig) (2)	**relaxed** (entspannt) (2)
d2 negative correlation	fending (abwehrend) (2)	agitated (aufgewühlt) (3)

General Procrustes Analysis generated as many dimensions as terms were available. For that reason, the number of dimensions in Free Choice Profiling was higher than using Fixed Terms, because the assessors tended to find more than twenty terms by Free Choice Profiling. With both Qualitative Behaviour Assessment methodologies, the main Dimensions 1 and 2 explained more than 10% of the variance. The Free Choice Profiling Dimension 1 explained 54.2% of variance; Dimension 2 explained 10.7% of the variance. Including even Dimension 3 (6.2%) 71.1% of variance is explained. The Fixed Term Dimension 1 of the individual scoring explained 51.7% of the variation. Dimension 2 of Fixed Term assessment the i-clips explained 13.2%. Including Dimension 3 71.9% of the variance is explained.

The analysis of the Free Choice Profiling data showed that the term “indifferent” (“gelangweilt”) was mentioned six times. The term “bored” (“gelangweilt”) was used by the observers six times to assess the i-clips with the Fixed Terms. The negative area of Dimension 1 was described with the term “requesting” (“auffordernd”) in both assessments. The positive area of Dimension 2 was described with the term “relaxed” (“entspannt”) with both assessment methods. This term was used two times. The generated terms of the Free Choice Profiling methods seemed to be similar to the Fixed Terms. It was noticeable that the negative area of Dimension 1 is described with exactly the same word - “requesting” (“auffordernd”); in the positive area of Dimension 2 the term “relaxed” (“entspannt”) loaded highest in both plots too. With both methods Dimension 1 could be summarised with the umbrella term “interest” (“Interesse”), while Dimension 2 illustrates “relaxation” (“Entspannung”).

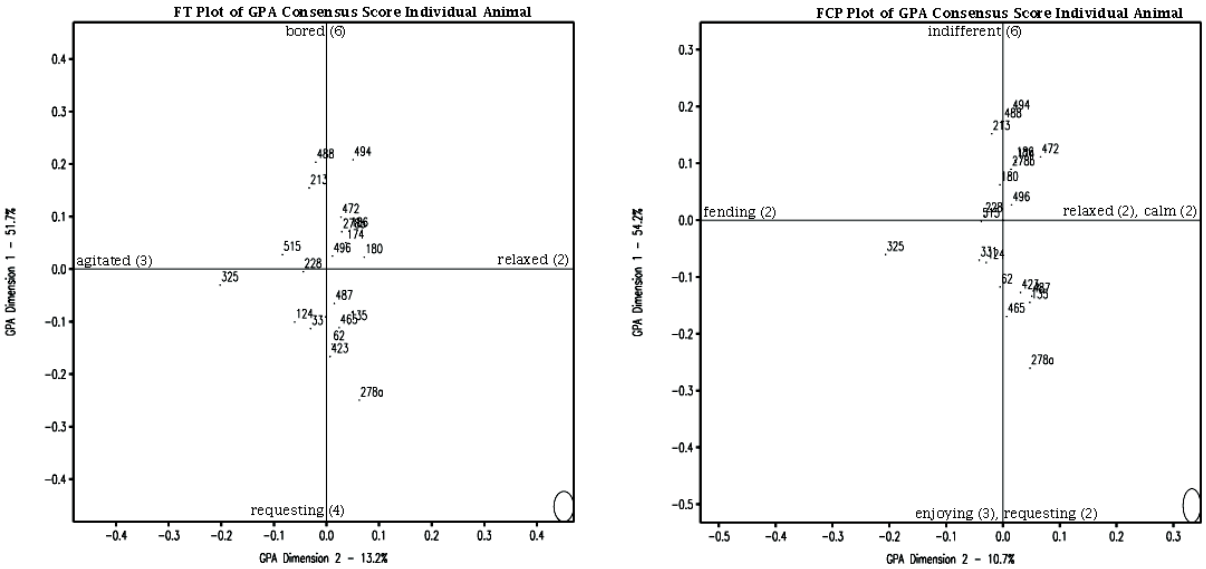


Figure 3: FT and FCP Score Plots for i-clips (n=20 clips)

The Fixed Terms Score Plot (Figure 3) illustrated a good differentiation between the clips on Dimension 1; but less differentiation was noticeable between descriptors of Dimension 2. Clip no. 325 was distinct because it was located near the negative correlation of Dimension 2, defined with the term “agitated” (“aufgewühlt”) - far away from the other clips. Only clip no. 278a was close to the negative correlation of Dimension 1. On the Free Choice Profiling Score Plot the clips are evenly distributed on the Dimension 1 between “indifferent” (“gleichgültig”) and “enjoying” (“genießend”). On Dimension 2 there is hardly a distribution recognizable. As for the Fixed Terms Score Plot, clip no. 325 tended to the negative correlation of Dimension 2, which was characterized with the term agitated (“aufgewühlt”). As on Fixed Terms Score Plot, clip no. 278a was located close to the end of Dimension 1 defined with the terms “enjoying“/“requesting” (“genießend”/“auffordernd”) - far away from the others.

Many clips were located on top of each other. This indicated a similar assessment of these clips. It was recognizable that nearly all clips had the same positions on both plots. The clips no. 213, 488 and 494 were located in the positive area of Dimension 1 on both consensus plots. Clip no. 496 was located near the centre on both plots. The clips no. 515 and 228 were very close to each other and located in nearly the same positions on both plots. The standard error circle encompassed clips no. 494 and 488 and clips no. 278b, 174 and 186 on the Fixed Terms Score Plot. On

Free Choice Profiling Score Plot the standard error encompassed clips no. 228 and 515 on the one hand and 187, 427 and 135 on the other hand.

A Pearson Correlation Coefficient was computed for both methods (Figures 4 and 5) for both dimensions to compare the scores. The correlation coefficient for the Dimension 1 (FCP vs. FT) had a value of 0.96 ($p>0.0001$). For Dimension 2, the correlation coefficient was slightly lower with a value of 0.88 ($p<0.0001$).

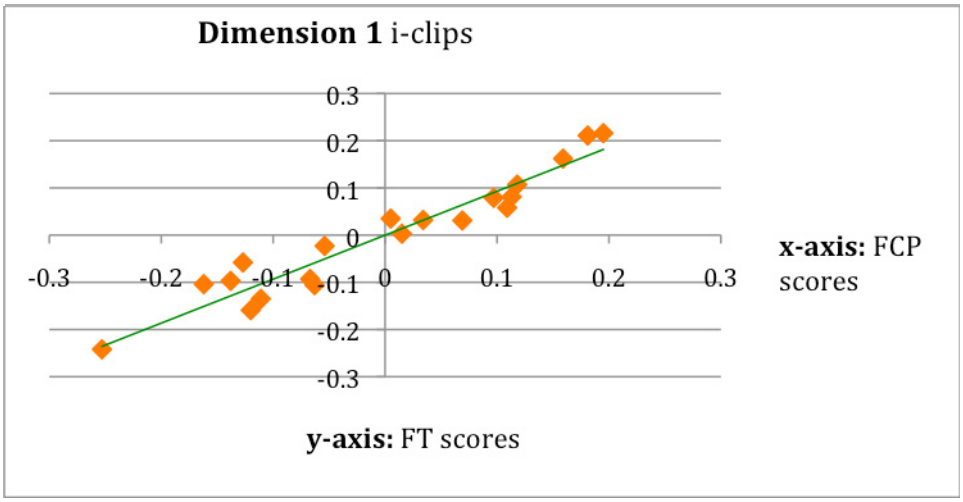


Figure 4: Scatter plot of Dimension 1 scores for i-clips originating from FCP (x-axis) and FT (y-axis) r=0.96 ($p<0.0001$)

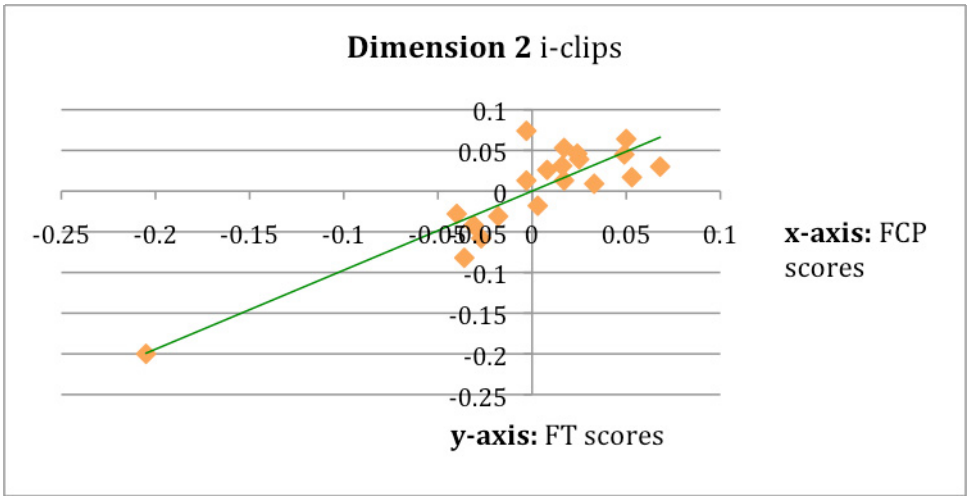


Figure 5: Scatter plot of Dimension 2 scores for i-clips originating from FCP (x-axis) and FT (y-axis) r=0.88 ($p<0.0001$)

3.2. Comparison of FCP and FT using Herd Clips

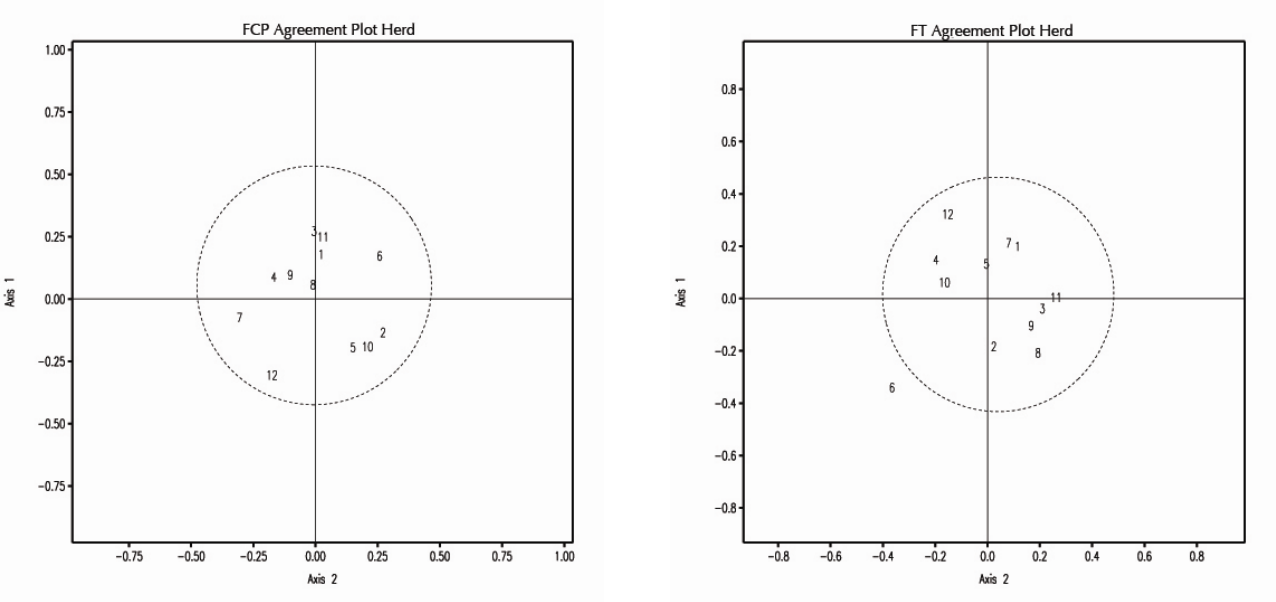


Figure 6: Comparison of FT and FCP observer agreements for herd assessment

The total percent variation explained by the consensus of the Free Choice Profiling was 65.40% of the total variation; the mean of 100 randomised profiles was 53.32%, the simulated variation had a value of 0.2116 (t_{99}, $p<0.001$). There was no outlier on the Free Choice Profiling Agreement Plot all twelve observers were located inside the confidence region. The one-sided t-test showed a significant difference between the total percent variation explained by the consensus and the simulated mean (df = 99, t= 26.3, p<0.001).

The Fixed Terms Agreement Plot of the h-clips showed one outlier, located outside the confidence region, which indicates a lower agreement. The Procrustes Statistic of Fixed Terms had a value of 66.18%. The mean of the randomised profiles was 51.48; the simulated variation had a value of 0.3163. The one-sided t-test showed significant differences between the total percent variation explained by the consensus and the simulated mean (df = 99, t = 26.2, p<0.001).

Table 4: Comparison of individual animal clips - FCP and FT Procrustes Statistic

Herd assessment		
GPA results	**FCP**	**FT**
Consensus profile	65.40%	66.18%
Randomised profile (mean ± variance)	53.32% ± 0.21%	51.48% ± 0.32%
t_{99} significance test	26.3($p<0.001$)	26.1($p<0.001$)

The dimensions of the h-clips were defined following the same principles as for the i-clips dimensions: most often mentioned terms with a 0.5 loading on the axes. Dimension 1 of both assessment methods of the herd assessments could be summarised with the umbrella term "relaxation" ("Entspannung"), while Dimension 2 can be called "interest" ("Interesse").

Table 5: Most frequently used terms with a loading of ≥ 0.5 in FCP and FT assessment of h-clips (number of times the term was used in brackets)

Dimension correlation	FCP	FT
d1 positive correlation	**relaxed** (entspannt) **(3),** **calm** (ruhig) **(2)**	**calm** (ruhig) **(5),** **relaxed** (entspannt) **(3)**
d1 negative correlation	tense (angespannt) (3)	distressed (gestresst) (4), agitated (aufgewühlt) (3)
d2 positive correlation	impatient (ungeduldig), busy (beschäftigt)	playful (verspielt)
d2 negative correlation	expectantly (erwartungsvoll) (2)	positively occupied (positiv beschäftigt) (2)

The dimensions created with the herd assessment are illustrated in Table 5. The positive correlation of Dimension 1 was characterized with the same terms in both assessment methods – "calm"/"relaxed" ("ruhig"/"entspannt"). The negative correlation of Dimension 1 was defined with "tense" ("angespannt") in Free Choice Profiling, which was used three times; the terms "distressed" (gestresst") and "agitated" ("aufgewühlt") characterise the negative correlation of dimension on with Fixed Terms. It was noticeable that there is no term mentioned many times to describe the positive correlation of Dimension 2 with both assessment methods. The presented terms "playful" ("verspielt"), "impatient" ("ungeduldig") and "busy" ("beschäftigt") were mentioned once. Free Choice Profiling Dimension 1 explained 39.6%, Dimension 2 10.1% and Dimension 3 explained 8.5% of the total variation. Together all three Free

Choice Profiling dimensions explained 58,2% of the variation. Dimension 1, 2 and 3 of the Fixed Term herd assessment explained in total 66% of the variation. Fixed Terms Dimension 1 explained 42%, Dimension 2 15.2% and Dimension 3 8.8%.

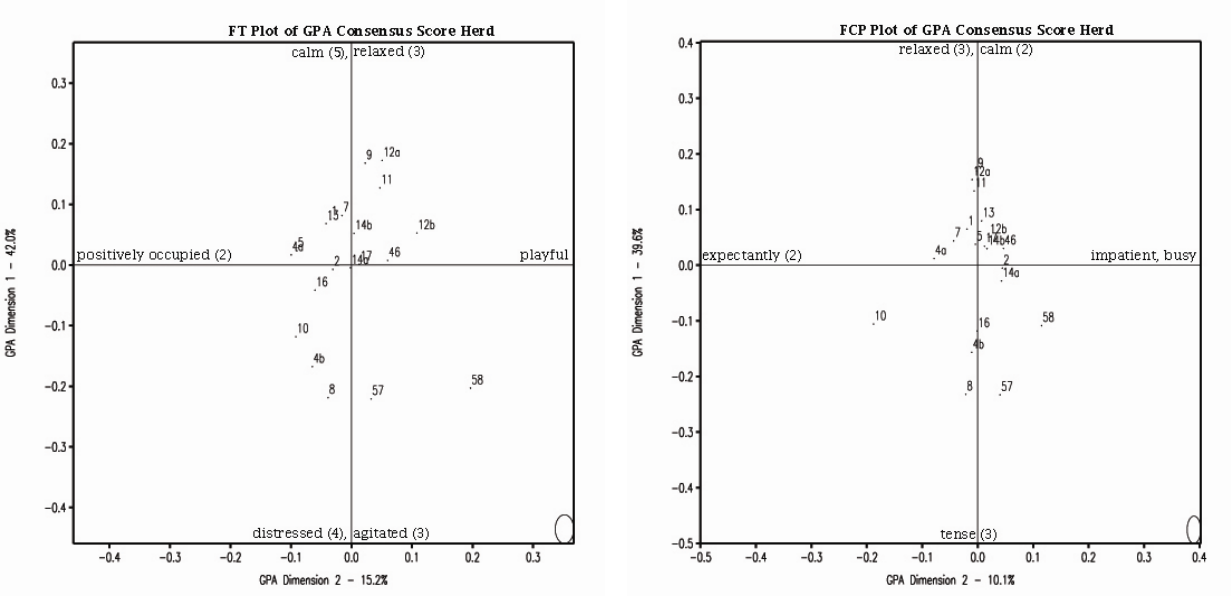


Figure 7: FT and FCP Score Plots of h-clips.

Both consensus Score Plots of both assessment methods illustrated a high distribution of the clips. There was a good distance between the clips. As explained with the i-clips, it is recognizable that some clips were located nearly on the same position on both Score Plots. This indicates that the observers tend to assess the video clips similar during both assessment methods. E.g. clips No. 8 and 57 were located on the same position. On the Fixed Terms Score Plot, both dimensions gave a good differentiation between the clips. There was hardly any clip located close to another clip, which indicated a high distribution. Clip no. 58 was located far away from the other clips. The standard error circle of the Fixed Terms Score Plot encompassed the clips no. 4a and 5, 17 and 14b in the centre and the clips no. 1, 7 and 13. All the other clips had a good distance to each other. It was noticeable that on the Free Choice Profiling plot the dimensions also created a good differentiation but most of the terms were closer to Dimension 1. Several clips were located on the top of each other near the centre. Clips no. 17, 14b and 12b were located inside the standard error circle as well as the clips no. 9, 12a and 11. Clip no. 10 was salient because it was located far away from the other clips close to the negative correlation of Dimension 2, which was defined with the umbrella term “interest” (“Interesse”),

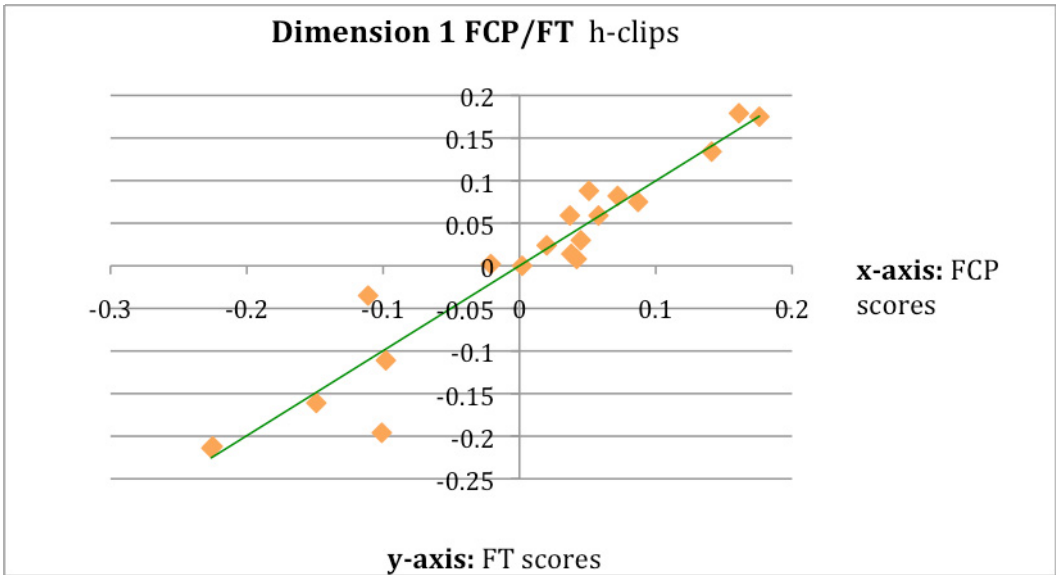


Figure 8: Scatter plot of Dimension 1 scores for h-clips originating from FCP (x-axis) and FT (y-axis) r=0.96 (p<0.0001)

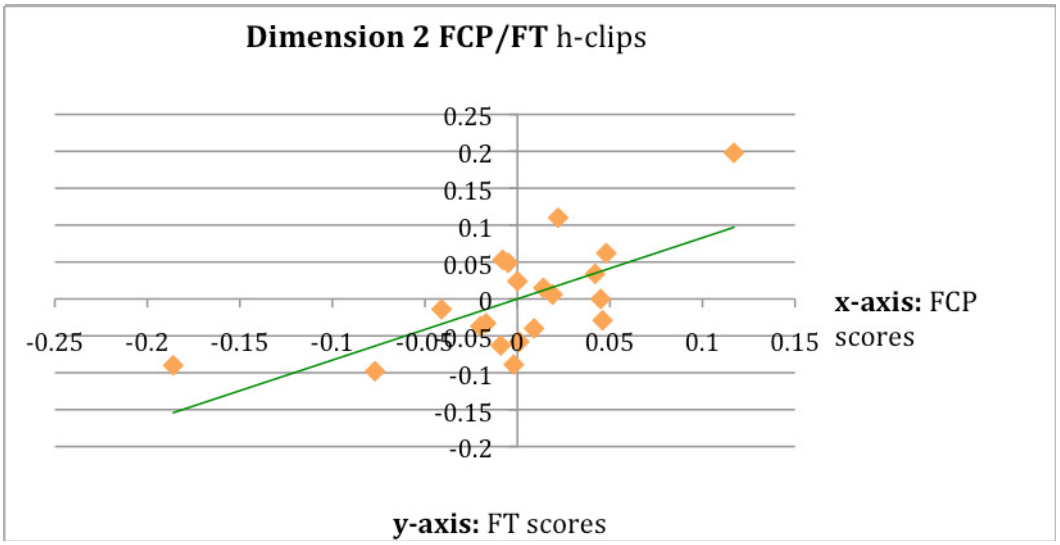


Figure 9: Scatter plot of Dimension 2 scores for h-clips originating from FCP (x-axis) and FT (y-axis) r=0.67 (p>0.001)

The Pearson Correlation Coefficient for Dimension 1 had a value of 0.96 (p<0.0001), which indicates a high correlation between the assessment methods. The Correlation Coefficient of Dimension 2 was lower with a value of 0.67 (p<0.001).

3.3. Comparison of Individual Animal and Herd Assessment

With Free Choice Profiling of individual animal and herd assessment assessors generated different numbers of terms, taking into account that some terms were generated often (Table 6). The number of the terms assessing the h-clips was higher. Free Choice Profiling of the h-clips gave 346 terms in total (including all same terms over all observers). The number of terms per observer ranged between 18 and 52 with a mean number of 29 terms (n=12 observers). Free Choice Profiling of the i-clips gave 206 terms in total; on average the assessors generated 17 adjectives. The number ranged from 10 to 32.

It was noticeable that the term “relaxed” (“entspannt”) was generated by all twelve observers during both assessment procedures. “Enjoying” (“genießend”) was found by all observers to describe the i-clips. All twelve observers chose the word “restless” (“unruhig”) to describe the h-clips. Eleven observers used the terms “calm” (“ruhig”), “curious” (“neugierig”) and “tense” (“angespannt”) to assess the h-clips. The term “indifferent” (“gleichgültig”) was generated to describe the individual clips by ten panellists. Nine assessors generated the word “bored” (“gelangweilt”) to evaluate the h-clips and the word “calm” (“ruhig”) to evaluate the i-clips. Seven panellists generated the term “distressed” (“gestresst”) to describe the h-clips, but only one assessor to describe the i-clips.

Table 6: Most often generated terms in FCP

Assessors	FCP h-clips	FCP i-clips
12/12	restless (unruhig) relaxed (entspannt)	relaxed (entspannt) enjoying (genießend)
11/12	calm (ruhig) curious (neugierig) tense (angespannt)	--
10/12	--	indifferent (gleichgültig)
9/12	bored (gelangweilt) attentive (aufmerksam)	calm (ruhig)
7/12	distressed (gestresst)	--

The dimensions of i-clips could be described with the umbrella terms “interest” (“Interesse”) and “relaxation (“Entspannung); the h-clip dimensions were described with the umbrella terms “relaxation” (“Entspannung”) and “interest” (“Interesse”). Although

Free Choice Profiling was conducted first, it is noticeable that panellists generated similar terms to describe different video clips.

Table 7: Dimensions of herd and individual animal assessment

	FCP	FT	Umbrella term
	i-clips		
D1	indifferent (+) enjoying/requesting (-)	bored (+) requesting (-)	**Interest**
D2	relaxed/calm (+) fending (-)	relaxed/enjoying (+) agitated (-)	**Relaxation**
	h-clips		
D1	relaxed/calm (+) tense (-)	calm/relaxed (+) distressed/agitated (-)	**Relaxation**
D2	impatient (+) expectantly (-)	playful (+) positively occupied (-)	**Interest**

The results showed a difference in the consensus profiles. Independent of the assessment method, the consensus profile of the herd assessments showed a higher agreement between observers than the consensus profile of individual animal assessments (Table 8). The total percent variation explained by the consensus of the individual animal assessment with Free Choice Profiling was 57.96%; Free Choice Profiling of the h-clips gave a consensus value of 65.40%. The total percent variation explained by the consensus of the Fixed Terms i-clips was 56.93. The assessment of the h-clips with Fixed Terms showed a consensus value of 66.18%.

Table 8: Summarised comparison of the consensus profiles i-clips and h-clips

Consensus profiles		
FCP Individual Animal		**FCP Herd**
57.96% ($p<0.001$)	**<**	65.40% ($p<0.001$)
FT Individual Animal		**FT Herd**
56.93% ($p<0.001$)	**<**	66.18% ($p<0.001$)

4. Discussion

The main aim of the present study was to investigate the comparability of Free Choice Profiling and Fix Terms assessment of dairy cows behaviour. The present study demonstrates similar results with both assessment methods but differences between individual and herd assessments. Free Choice Profiling as well as Fixed Terms assessment showed approximately the same explained total variation; this is an important result regarding the main question of the present study. Observers often generated the same terms with Free Choice Profiling as given with the Fixed Terms. These results show that on-farm Qualitative Behaviour Assessment of herds may be conducted with the developed Welfare Quality standard protocol for dairy cattle, if the complex Free Choice Profiling is not possible.

As compared with several previous studies (Wemelsfelder et al., 2000, 2001a; 2001b; Rousing and Wemelsfelder, 2006), the consensus profiles showed only moderate observer agreement. Procrustes Statistic was nearly 81% in Wemelsfelder et al. (2000), while the consensus profile ranged between 66.4% and 71.4% in the study of Rousing and Hunter (2006) and between 72.6% and 85.3% in the study of Wemelsfelder et al. (2001a). However, the present study did not show a difference between Fixed Terms and Free Choice Profiling for both the assessment of individual animals as well as the assessment of groups of cows.

Generally, the presented study revealed a certain difference between individual assessment and the herd assessment. Herd assessment revealed higher agreement between panellists as given by the Procrustes Statistics. Three terms contained in the Welfare Quality standard protocol were changed to presumably better describe expressive quality of behaviour of individual animals in a social licking interaction. It is recommendable to use terms that correspond to the specific situation. Therefore, the terms “enjoying” (“genießend”), “requesting” (“auffordernd”) and “intrusive” (“aufdringlich”) were chosen. Results may be different with other terms, but the generated Free Choice Profiling-terms of the observers confirmed this choice. On the other hand, the quality of the individual animal clips was moderate and the clips were in black and white. For this reason, it might have been difficult to accurately identify the behaviour of the cows. Some cows were lying nearly motionless. The emotional state of these animals could have been more difficult to detect. These facts may explain

the lower consensus profiles of the present study and indicate the importance of using video clips with good quality in further studies.

On the first rating day Free Choice Profiling of twenty i-clips and twenty h-clips were done in two sessions. One week later the Fixed Terms assessment was also done in two sessions of twenty clips each. After each session there was a break. In total, forty video clips were assessed on one rating day. Watching forty clips and scoring many terms very concentrated might be exhausting. This fact may lead to comparably low consensus profiles. The high variance was intentional to avoid monotony and comparison with other video clips seen before. The video clips were chosen to show a high diversity of cow behaviour. As a consequence, the generated terms should reflect this behaviour diversity. Panellists should assess every single video clip as independently as possible. For that reason, the clip sequence was chosen to avoid similar clips sequenced. As a result it should be possible to assess the actual seen clip without remembering the clips seen before and comparing with previously seen clips.

The assessment procedures started with the Free Choice Profiling to ensure that the Fixed Terms of the Welfare Quality protocol did not influence observers. Therefore, the observers did not see this protocol before the procedures started. After the panellists had created their terms, those with close meanings were excluded and not entered in the Free Choice Profiling rating lists. As stated in Wemelsfelder et al. (2000), different terms often have close meanings, e. g. “request”/“demand” (“auffordern”/”fordern”) or “enjoying”/”appreciative” (“genießend”/”genüsslich”). But even if terms tone very similar, the observer personally could have had different contents in mind. But it is also necessary to mention that similar words were chosen to evaluate animal’s behaviour, which do not have the same meanings (Wemelsfelder et al., 2000). For that reason, it may be advantageous if no terms are removed.

The observer group consisted of students and professors. Thus, there was a variance in the knowledge of dairy cattle behaviour. Furthermore, there were four persons well trained in Free Choice Profiling while the other eight persons were laypeople. Some assessors generated a large number of terms. For the herd assessment, Assessor 5, a woman, generated 52 terms. This high number of generated terms might be the effect of unspecific introduction, high creativity or by experience in Qualitative Behaviour Assessment. On the other hand, Assessor 2, a man, only found eighteen words, which might indicate that Assessor 2 did not have experience

in Qualitative Behaviour Assessment or that women might be more creative than men. All four men found an average of 11.25 terms while the eight women found 25.88 terms, but it is not heeded that some terms are generated more often. All four men generated the terms “relaxed” (“entspannt”) and “calm” (“ruhig”). Three of them generated the adjectives “enjoying” (“genießend”) and “tense” (“angespannt”). In contrast there is no term generated by all eight women. Seven women found the adjective “relaxed” (“entspannt”), six women the word “indifferent” (“gleichgültig”) and five women the term “calm” (“ruhig”). But in total, women generated more terms, which might indicate that women need more terms to describe all their discernible dimensions of cow’s behaviour or that women might be more creative than men in this context. This outcome could be a subject for further research.

Because of the fact that some creative panellists needed a lot of time to tick the scale of every term it is recommendable to give a time limit in the term-generating-phases of Free Choice Profiling as done in studies done by Wemelsfelder et al. (2009c). This may avoid generating many terms by very creative panellists. This would simplify the scoring and may reduce time consuming scorings and spare concentration. Furthermore, the panellists would be forced to write down the adjectives that came first into their mind first.

Although the Fixed Terms assessment was conducted one week after the Free Choice Profiling, most clips revealed similar scores. E.g. i-clips no. 325 and 278a are located close to the negative end of Dimension 1 far away from the other clips on Fixed Terms plot as well as on Free Choice Profiling plot. This fact confirms the observer agreement of the assessments.

On the herd consensus Scores Plot clip no. 58 was salient. On the Fixed Term plot it is located far away from the other clips. Clip no. 10 on the herd Free Choice Profiling consensus Score Plot is located close to the negative end of Dimension 2, which is defined as “expectantly” (“erwartungsvoll”). All in all the h-clips are highly distributed on the consensus Score Plot - the dimensions are worked out clearly. This effect is only given with the h-clips. The high distribution of the h-clips on the Score Plots may have been caused by the high diversity of behaviour that could be recognized by assessing a herd. Because of the fact that a herd offers more different behaviour expressions it might be easier to find a high clip-variability for assessments. Therefore, it might be possible to cover further more behaviour dimensions.

A discussion after assessing the i-clips with Fixed Terms revealed that panellists had problems to assess an individual's behaviour with given terms. They indicated that they were missing some terms to describe the behaviour as seen. Otherwise they explained that there were terms they did not need to evaluate the social licking interaction. The debriefing also showed that panellists hardly had problems finding terms to describe seen behaviour of the herds. A group of animals offers a lot of different behaviours. Therefore, it was not difficult to find several descriptors. Fixed Terms assessment was conducted almost without any problems. This might be the most important difference to the individual assessment in a special interaction situation. Although the agreements were very similar with both assessment methods, the panellists indicated that Free Choice Profiling, especially for the i-clip assessment, gave the possibility to score independently. With the Fixed Terms, they felt restricted in describing animal behaviour, especially if the individual animal should be assessed in special interactions. This may be a result of conducting Free Choice Profiling first.

Furthermore, the Fixed Term Agreement Plot of the individual assessment showed two outliers. Both were women that created far more terms than the other panellists. This may indicate that Fixed Terms are unsuitable for creative panellists. Because of the limited number of available terms they are not able to assess all dimensions of cow's behaviour they perceive. This fact could be subject to further research.

5. Conclusion

In the present study, video clips of dairy cattle were evaluated using Free Choice Profiling and Fixed Terms, respectively, for Qualitative Behaviour Assessment. The results show that Fixed Terms can be used for on-farm assessment of a dairy cattle herd, because the Free Choice Profiling and Fixed Terms results are very similar: the observer agreement as well as the dimensions' characterization exhibit very similar results. Panellists were able to generate exactly the same terms as developed by experts for the Fixed Terms. Although Free Choice Profiling exhibits advantages (e.g.: independence of the panellists) it is possible to use Fixed Terms if using Free Choice Profiling is not possible due to the time and personal effort. Whether twenty Fixed Terms, as suggested in the Welfare Quality protocol, are enough for panellists to assess all dimensions of behaviour, or whether there is an influence of the emotional state of the observers on their evaluations, or whether there are gender differences in the assessment are questions that could be subjects of further research.

However, it remains open if Fixed Terms protocols work equally well for Qualitative Behaviour Assessment in pigs and poultry as proposed in the Welfare Quality assessment protocols. Further research in this area is therefore recommended.

6. Summary

Qualitative Behaviour Assessment is used to describe animal welfare with terms such as “relaxed” (“entspannt”) or “enjoying” (“genießend”). The present study investigated if given Fixed Terms can be used to evaluate dairy cattle behaviour instead of time-consuming Free Choice Profiling. Additionally, the study should reveal similarities or differences in assessing individual animals in a specific social licking interaction or groups of animals with Free Choice Profiling and Fixed Terms.

Therefore, twenty video clips of a herd (h-clips) and twenty individual animal clips (i-clips) were assessed by twelve observers - eight women and four men. For Free Choice Profiling, a term-generating session was prepared, where the assessors generated their own adjectives that were entered into the Free Choice Profiling protocol. The Welfare Quality standard protocol served as Fixed Terms for the herd assessment. This protocol was adapted for individual animal assessment with the Fixed Terms to cover behaviours seen in specific social licking situation. The rating started with Free Choice Profiling. One week later the rating was repeated using Fixed Terms.

The statistical analysis was conducted with Generalised Procrustes Analysis. The Procrustes Statistics of the i-clips with Free Choice Profiling had a value of 57.96%; with Fixed Terms 56.93%. With both methods, the panellists often described the video clips with the same terms – “requesting” (“auffordernd”) and “relaxed” (“entspannt”). If there were different words, they often had an equal meaning – “fending” (“abwehrend”) and “agitated” (“aufgewühlt”). The Procrustes Statistics of the h-clips had a value of 65.40% with Free Choice Profiling and 66.18% with Fixed Terms. The assessors described the clips with the same words – “relaxed” (“entspannt”) and “calm” (“ruhig”). These results reveal that Free Choice Profiling of herds gave a higher agreement as Free Choice Profiling of individual animals.

The study showed that Fixed Terms could be used for on-farm assessment of dairy cattle if Free Choice Profiling is not possible because of the time and personal effort.

7. Zusammenfassung

Qualitative Verhaltensbeurteilung beschreibt Tierwohlbefinden mit Adjektiven wie „entspannt“ oder „genießend“. Ziel der Arbeit war es, zu untersuchen, ob anstelle des zeitaufwendigen Free Choice Profiling (FCP) auch fix vorgegebene Begriffe zur on-Farm-Beurteilung von Milchviehverhalten geeignet sind. Weiters soll die Studie Gemeinsamkeiten und Unterschiede der Einzeltier- und Herdenbeurteilung mittels FCP und fixen Begriffen aufzeigen, die Einzeltiere wurden beim sozialen Lecken beurteilt. 20 Herdenclips (h-clips) und 20 Einzeltierclips (i-clips) wurden von 12 Beobachtern (8 Frauen, 4 Männer) beurteilt. Um FCP durchführen zu können, generierten die Beobachter vorweg ihre eigenen Begriffe, welche anschließend in das Beurteilungsprotokoll eingetragen wurden. Für die Beurteilung mit fixen Begriffen wurde für die Herde das Welfare Quality Standard Protokoll verwendet, welches für die Beurteilung der Einzeltiere angepasst wurde, um die Verhaltensweisen des sozialen Leckens zu erfassen. Zuerst wurde FCP durchgeführt; eine Woche später die Beurteilung mit den fixen Begriffen. Die Datenauswertung erfolgte mittels einer Generalised Procrustes Analyse. Die Beobachterübereinstimmung der Einzeltierbeurteilungen ergab einen Wert von 57,96% mit FCP; 56,93% mit den fixen Begriffen. Sowohl mit FCP als auch mit den fixen Begriffen wurde die Clips mit denselben Begriffen beurteilt („auffordernd“ und „entspannt“). Gab es unterschiedliche Begriffe, so war dennoch ihre innere Bedeutung sehr ähnlich („abwehrend“ und „aufgewühlt“). Die Beobachterübereinstimmung der Herdenbeurteilungen ergab mit FCP einen Wert von 65,40%; mit den fixen Begriffen 66,18%. Auch hier wurden die Clips mit denselben Worten beschrieben („entspannt“ und „ruhig“). Die Herdenbeurteilungen erzielten eine höhere Übereinstimmung. Die Studie zeigt, dass fixe Begriffe für on-Farm-Beurteilungen von Milchkühen eingesetzt werden können, wenn FCP aufgrund des hohen Zeit- und Personalaufwandes nicht möglich sein.

8. Appendix

Table 9: Verbal description of the h-clips.

Clip	Verbal description of the herd clips
14b	Herd of cows in a free stall barn with straw litter
57	Herd of cows in an outdoor area. A few animals are located at the watering place.
4a	Holstein Friesian cows located at the feedlot and lying in the boxes
10	Cows are located at the canopied outside area.
8	Holstein Friesian cows in a free stall barn.
4b	A herd of cows staying outside in the sun.
14a	A herd of cows. Nearly all cows rest in the litter.
12b	Three cows lying; one cow standing in a littered lying area.
7	German Simmental cows in a free stall barn. A certain degree of movement is recognizable. Some cows walk around, other individuals are resting in the littered boxes.
11	Holstein Friesian cows in a littered free stall barn. Two cows in the middle of the screen near the wood column are comparatively active.
16	Large group of Holstein Friesian cows. Most animals rest in the moderate littered boxes. One cow is located at the watering place.
12a	Many cows at the feedlot. One cow tries to find a feeding place.
2	German Simmental cows; slatted floor is well recognizable. The boxes are moderate littered. One cow leaves the automatic feeder.
9	Holstein Friesian cows. Three lying cows are well recognizable as well as the slatted floor. One cow, located at the left screen, looks at the camera.
1	Hindquarters of German Simmental cows in the front and the back of the screen; a very brown cow with a white head can be seen totally.
46	Herd of cows different breeds; three cows at the feedlot; behind a Holstein Friesian cows walking.
58	Cows at the feedlot at feed intake.
5	Very wet and dirty floor is well recognizable. One cow stands on the lying boxes with its forelegs.
17	Cows at the feedlot; a very grey is being licked by another cow.
13	Cows staying at the outside area in the sun. Three cows are located directly at the feedlot.

Table 10: Verbal description of the i-clips.

Clip	Verbal description of the individual clips (social licking)
62	Actor and receiver are located in the front.
124	Two standing cows practice social licking.
135	In the middle of the herd the actor is lying while the receiver, a very white cow, is standing.
174_176	Clip is masked on the left side. The actor and the receiver are located in the left part of the screen.
180	Actor and receiver are lying in the middle of the pictures.
186	Actor stands while the receiver lies.
213	Actor and receiver are located in the left corner; the moderate clip is masked in the front to hide the cows at the feedlot.
228	Receiver is lying while the actor stands in front of him. At first there is no licking cognizable.
278a	Actor and receiver stay in the front of the pictures directly at the feedlot.
278b	Receiver and actor are situated in the left corner.
325	Actor stands with Receiver near the entrance. The receiver moves the head frequently.
331	Actor and receiver are standing in the front at the feedlot. The receiver shakes the head.
423	Actor lies behind the standing receiver that raises its head up and down.
465	Actor and receiver stand in the front at the feedlot; left eye of the receiver is well cognizable.
472	Actor and the very white receiver lie in the right corner.
487	Actor and receiver are located in the background.
488	Receiver and actor lie in the right corner.
494	Actor lies behind the well cognizable receiver. The receiver keeps calm all time. The quality of the clip is moderate.
496	Actor lies behind the well cognizable receiver. The quality of the clip is moderate.
515	Actor and receiver stand at the feedlot.

FT Scoring Protocol – Individual Animal Clips

Min. Max.

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.

Genießend

	Min.	Max.
Lebhaft		
	Min.	Max.
Auffordernd		
	Min.	Max.
Irritierbar		
	Min.	Max.
Unbehaglich		
	Min.	Max.
Aufdringlich		
	Min.	Max.
Apathisch		
	Min.	Max.
Glücklich		
	Min.	Max.
Gestresst		

FT Scoring Protocol – Herd Clips

Min. Max.

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.

Genießend

	Min.	Max.
Lebhaft		
Auffordernd		
Irritierbar		
Unbehaglich		
Aufdringlich		
Apathisch		
Glücklich		
Gestresst		

FCP terms created for assessment of the h-clips

Assessor 1 (woman)	Aktiv, angespannt, defensiv, suchend, erwartungsvoll, drohend, ungeduldig, neugierig, erschöpft, behaglich, aggressive, ruhig, vorsichtig, unaufgeregt, freundlich, ängstlich, erstarrt, aufmerksam, gemütlich unentschlossen, desinteressiert, unruhig, entspannt
Assessor 2 (man)	Durchetwasgestresst, aufmerksam, abwartend, kämpferisch, beengt, misstrauisch, suchend, entspannt, abgestumpft, genüsslich, gleichgültig, stur, neugierig, gelangweilt, angespannt, wartend, unruhig
Assessor 3 (man)	Angespannt, genervt, neugierig, passiv, unaktiv, aufmerksam, abwehrend, scheu, imeinklang, gestresst, gierig, ängstlich, gleichgültig, gereizt, unzufrieden, entspannt, aggressive, interessiert, geschreckt, unterfordert, ruhig, vorsichtig, unruhig, stürmisch, genießend, überfordert, gelangweilt
Assessor 4 (woman)	Einverstanden, aufmüpflig, genervt, hektisch, entspannt, geladen, sicher, vorlaut, übellaunig, angespannt, abgekehrt, gierig, gequält, ruhig, k.o., insichgekehrt, interessiert, rüpelhaft, extrovertiert, ausgeschlafen, bedrängt, fit, aufgewühlt, aktiv, resigniert, geschäftig, freundlich, selbstbewusst, ungeduldig, unhöflich, unterfordert, unruhig, stressfrei, tolerant, aufmerksam, unsicher, vertraut, zurückgezogen, überfordert, unentschlossen, unfrei, zufrieden, beengt, gelangweilt, desinteressiert, ungezwungen, unternehmungslustig, gemütlich, unausgeglichen
Assessor 5 (woman)	Alarmiert, orientierungslos, ruhig, belästigend, entspannt, träge, gebeutelt, sozial, lustlos, ungestüm, erwartend, fixiert, unentschlossen, wütend, angespannt, betrübt, sichbeengtfühlend, sehnsüchtig, neugierig, perspektivlos, enttäuscht, geborgen, unfreundlich unabhängig, ungeduldig, genießend, gestresst, fehlamplatzfühlend, beschäftigt, asozial, inpositi-

	vererwartung, sichunwohlfühlend, suchend, unbekümmert, angewidert, aggressiv, aktiv, ausweichend, sichungemütlichfühlend, überlastet, geruhsam genervt, unterdrückt, zielgerichtet, unzufrieden, angriffslustig, sichwohlfühlend, unruhig, gleichgültig, frustriert, zögerlich, gelangweilt
Assessor 6 (woman)	Entspannt, lustlos, unausgeglichen, neugierig, unruhig, frustriert, erschöpft, abwesend, genervt, gleichgültig, motiviert, unterfordert, angespannt, gelassen, erwartungsvoll, gestresst, passiv, zufrieden, unentschlossen, interessiert, gelangweilt
Assessor 7 (woman)	Bedrängt, entspannt, gelangweilt, nervös, genüsslich, aktiv, sozial, unruhig, verzweifelt, passiv, dominierend, angespannt, aufmerksam, abwehrend, irritiert, ruhig, ungewiss, unsicher, neugierig, leidend, aufgewühlt, erkundend, resignierend, überfordert, unwohl, erduldend, gestresst, unfreundlich interessiert, unglücklich, freundlich
Assessor 8 (woman)	Unsicher, belästigt, gelangweilt, freundlich angespannt, grob, interessiert, unmotiviert, sicher, neugierig, teilnahmslos, ängstlich, gestresst, gesellig, unruhig, wohl, konaktsuchend, aufdringlich, entspannt, herrschend, nervös, unfreundlich, gleichgültig, ruhig, unwohl, zufrieden, furchtlos, aufmerksam, kontaktfreudig, aktiv, aggressiv
Assessor 9 (woman)	Beengt, beobachtend, dominant, aufgescheucht, interessiert, rastlos, bestimmend, nervös, sichpräsentierend, ruhig, entspannt, gereizt, aufgewühlt, analysierend, einfordernd, gleichgültig, friedlich bedrängt, genießend, lebhaft, herausfordernd, ungestört, zielstrebig, uneinig, unruhig, neugierig, unentschlossen, rechthaberisch, unwohl
Assessor 10 (woman)	Angespannt, sichbedrängtfühlend, passiv, interessiert, dominierend, aktiv, neugierig, unzufrieden, ratlos, aufgeregt, aggressiv, fordernd, ruhig, unwohlsein, entspannt, angrifflus-

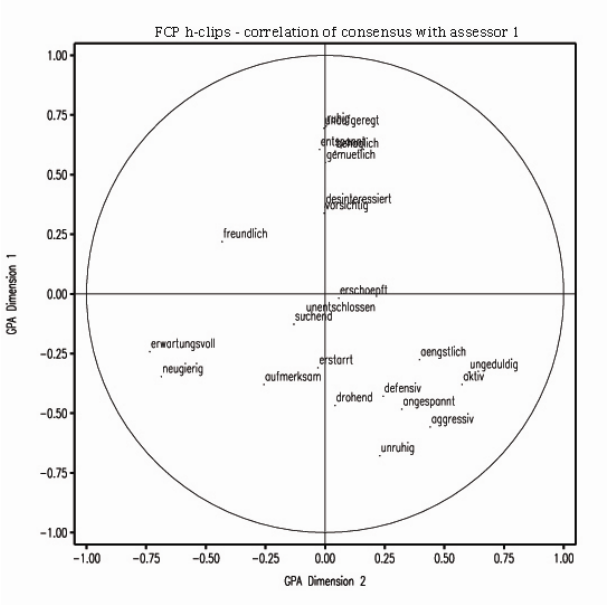
	tig, aufmerksam, gestresst, streitlustig, ängstlich, gelangweilt, unruhig, sichbeengtfühlend
Assessor 11 (man)	Abgelenkt, angespannt, genervt, passive, unsicher, gestresst, drohend, aufetwaswartend, aggressiv, entspannt, unruhig, stürmisch, zufrieden, aktiv, aufmerksam, ruhig, gelangweilt, neugierig
Assessor 12 (man)	Aktiv, entspannt, irritiert, gesellig, ablehnend, aufmerksam, unentschlossen, unruhig, neugierig, frustriert, aktiviert, teilnahmslos, unwohl, vertrauensvoll, unfreundlich genießend, ruhig, aggressiv, angespannt, gleichgültig, unsicher, fremdbestimmt, ungeduldig, genervt

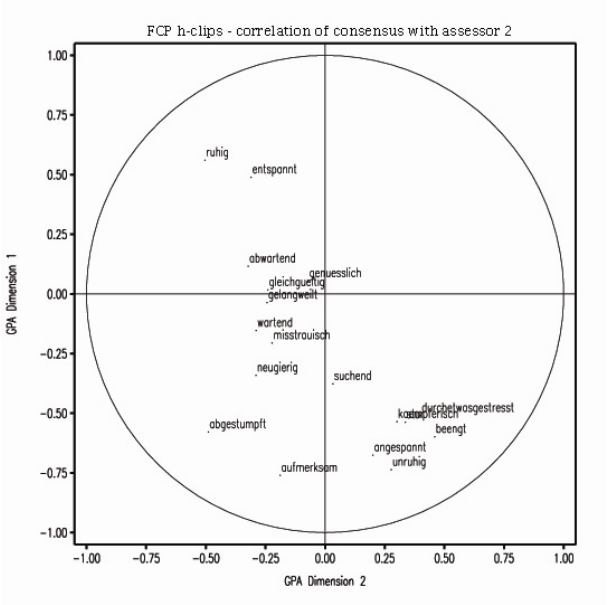
FCP terms created for assessment of the i-clips

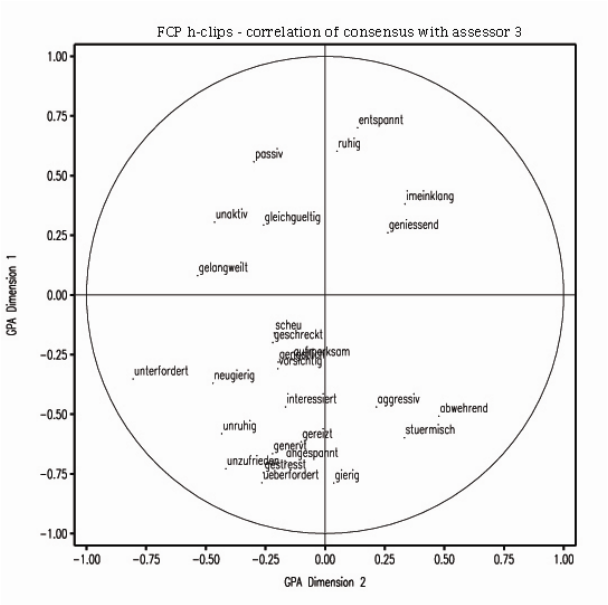
Assessor 1 (woman)	Träge, entspannt, duldend, genießerisch, lässt sich aktiv belecken, aufmerksam, tief entspannt, gleichgültig, sanft fordernd,
Assessor 2 (man)	Gleichgültig, entspannt, ruhig, unterlegen, abwartend, stumpf, unsicher, auffordernd, freundlich genüsslich
Assessor 3 (man)	Ruhig, entspannt, angespannt, genüsslich, phlegmatisch, uninteressiert, abwehrend, gleichgültig, duldend, zufrieden
Assessor 4 (woman)	Desinteressiert, entspannt, gestört, genießend, wohlig, ruhig, bedrängt, unbefriedigt, friedlich, erregt, aufmunternd, dringlich, neutral, müde, überlegen, zugetan, bittend, wehmütig, zustimmend, empört, verwundert, hat's nötig, gelangweilt, zulassend, dirigierend, wohlwollend, beengt, unterdruck, ungeduldig, wohltuend, fordernd, unzufrieden
Assessor 5 (woman)	Abwartend, relaxed, ambivalent, unsicher, hinspürend, festgefroren, duldend, überrascht, unentschlossen, prüfend, genießend, fordernd, zustimmend, mit sich selbst beschäftigt, phlegmatisch, ruhend, interessiert, sich hingebend, erwartend, statisch, gleichgültig, unbeeindruckt, ertragend, unbe-

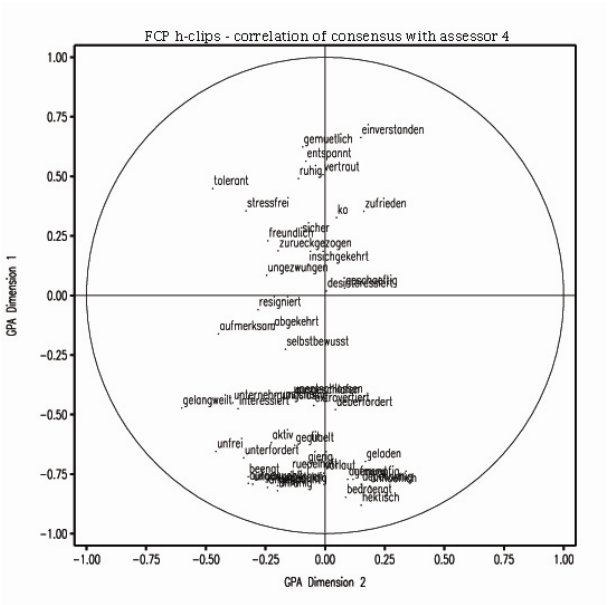
	geistert, anständig, widerwillig, vertrauensvoll, bereitwillig, kooperativ, mitmachend, liebesbedürftig
Assessor 6 (woman)	Gleichgültig, desinteressiert, entspannt, genüsslich, zufrieden, hingebungsvoll, gelassen, neutral, auffordernd, interessiert, passiv, abwägend, gelangweilt, verunsichert, überrascht, unentschlossen, animiert
Assessor 7 (woman)	Gleichgültig, genüsslich, stoisch, passiv, abwehrend, ruhig, erduldend, gelangweilt, hingebungsvoll, traurig, überfordert, auffordernd, niedergeschlagen, aktiv, selbstbewusst, aufmerksam, freundlich, entspannt, nett, beruhigend
Assessor 8 (woman)	Ruhig, angespannt, genießend, zufrieden, desorientiert, unwohl, entspannt, aufgeregt, genervt, fordernd, unbeteiligt, gleichgültig, kontaktfreudig, freundlich, suchend, aktiv, erfreut, anteilnehmend, gestresst
Assessor 9 (woman)	Ruhig, entspannt, gleichgültig, genießend, wohlwollend, angewurzelt, reaktionslos, ignorant, einfordernd, erwartungsvoll, hoffnungsvoll, tolerierend, regungslos, erfreut, willig
Assessor 10 (woman)	Entspannt, ruhig, fadisiert, desinteressiert, antriebslos, genießerisch, lustvoll, freudig, aufgeweckt, passiv, unbeeindruckt, angespannt, fordernd, zufrieden, aktiv, mehrwollend, erregt, sich anbietend
Assessor 11 (man)	Ruhig, passiv, gleichgültig, entspannt, zufrieden, genießend, aufmerksam, fordernd, interessiert, angespannt
Assessor 12 (man)	Gleichgültig, entspannt, ruhig, genüsslich, auffordernd, entrückt, selig, teilnahmslos, angespannt, frustriert, annehmend, schläfrig, neutral, aktiv, interaktiv

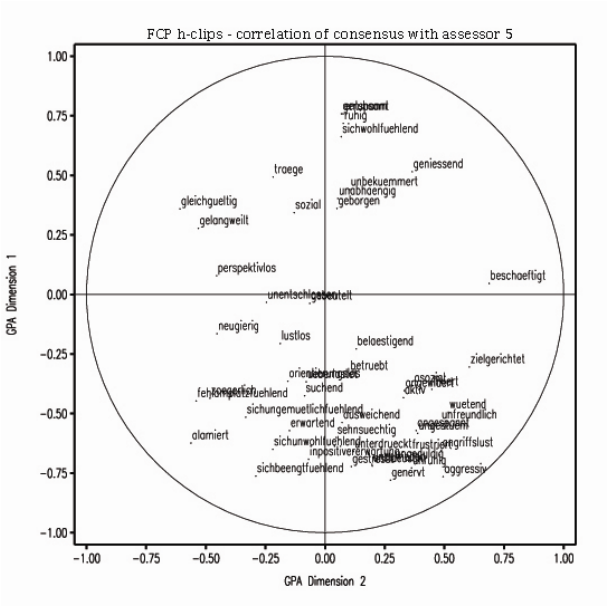
FCP Word Charts of 12 observers for h-clips

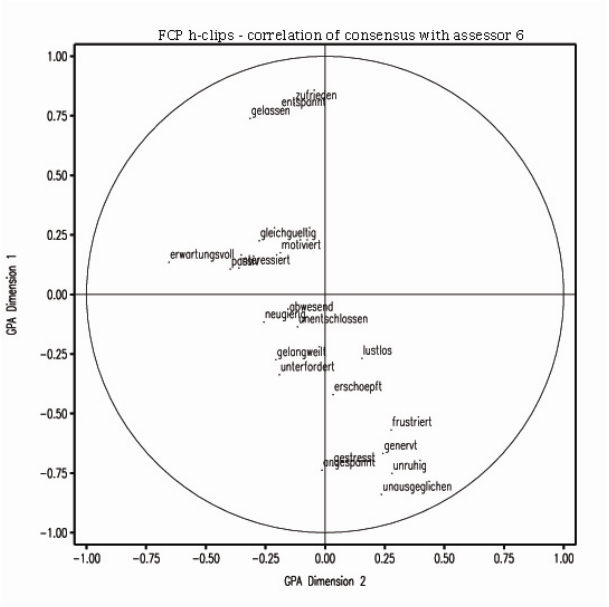


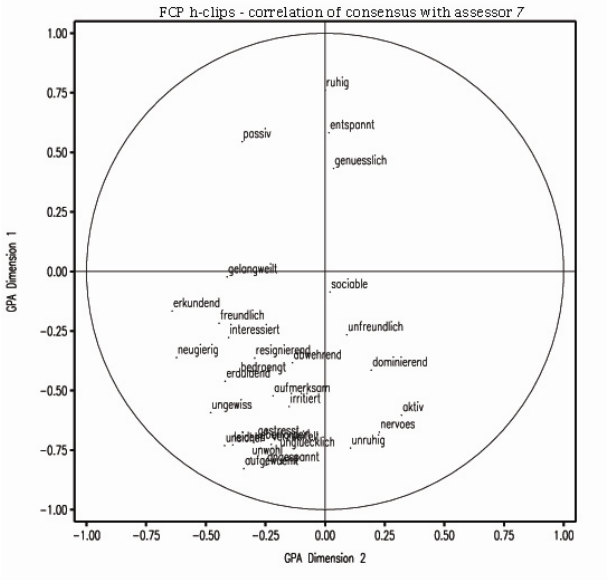


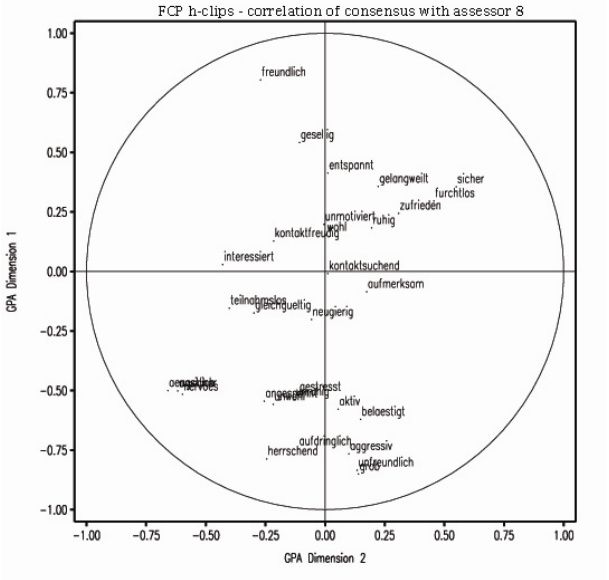


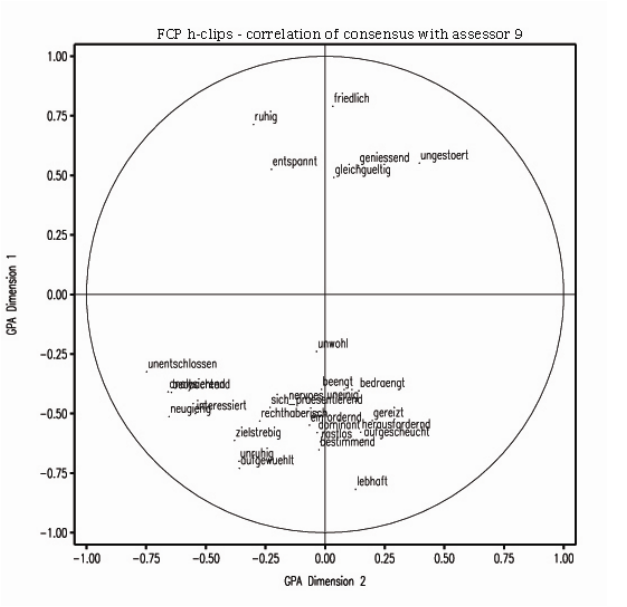


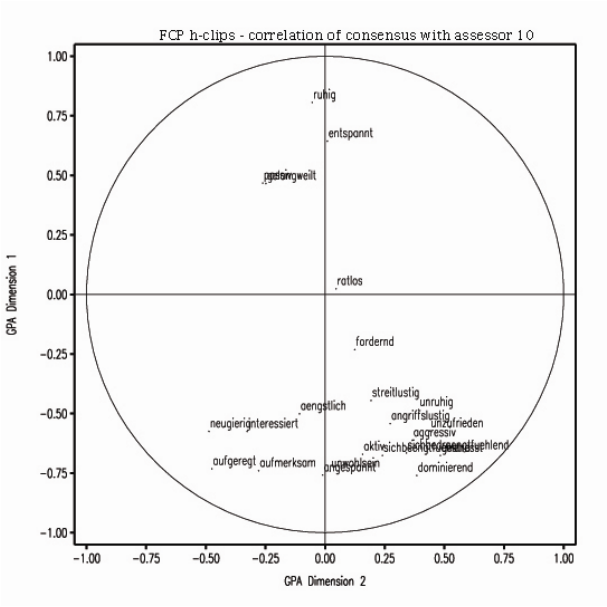

FCP h-clips - correlation of consensus with assessor 5
GPA Dimension 1
GPA Dimension 2
ruhig
sichwohlfuehlend
geniessend
traege
unbekuemmert
unabhaengig
geborgen
gleichgueltig
sozial
gelangweilt
perspektivlos
beschaeftigt
neugierig
lustlos
beloestigend
zielgerichtet
betruebt
suchend
aktiv
wuetend
unfreundlich
sichungemuetlichfuehlend
ausweichend
erwartend
sehnsuechtig
alarmiert
sichunwohlfuehlend
angriffslust
sichbeengtfuehlend
genervt
aggressiv

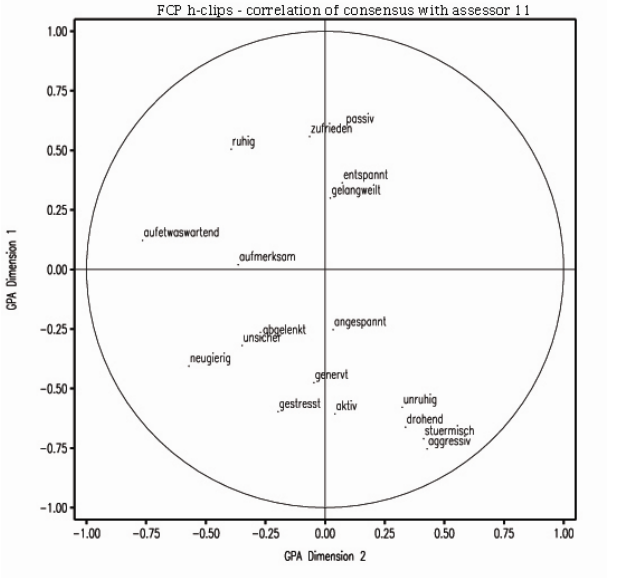

FCP h-clips - correlation of consensus with assessor 6
GPA Dimension 1
GPA Dimension 2
zufrieden
entspannt
gelassen
gleichgueltig
motiviert
erwartungsvoll
interessiert
obwesend
neugierig
entschlossen
gelangweilt
lustlos
unterfordert
erschoepft
frustriert
genervt
gestresst
unruhig
unausgeglichen

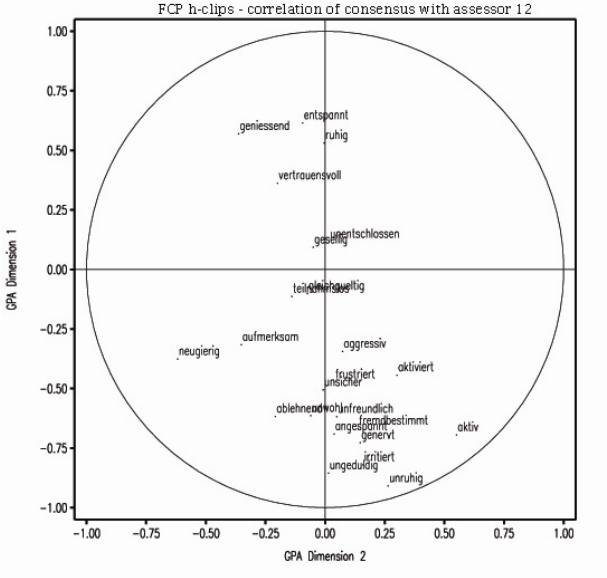

FCP h-clips - correlation of consensus with assessor 7
GPA Dimension 1
GPA Dimension 2
ruhig
possiv
entspannt
genuesslich
gelangweilt
sociable
erkundend
freundlich
interessiert
unfreundlich
neugierig
resignierend
abwehrend
dominierend
aufmerksam
irritiert
ungewiss
oktiv
nervoes
unruhig


FCP h-clips - correlation of consensus with assessor 8
GPA Dimension 1
GPA Dimension 2
freundlich
geselig
entspannt
gelangweilt
sicher
furchtlos
zufrieden
unmotiviert
ruhig
kontaktfreudig
interessiert
kontaktsuchend
aufmerksam
neugierig
gestresst
oktiv
beloestigt
aufdringlich
aggressiv
herrschend
unfreundlich

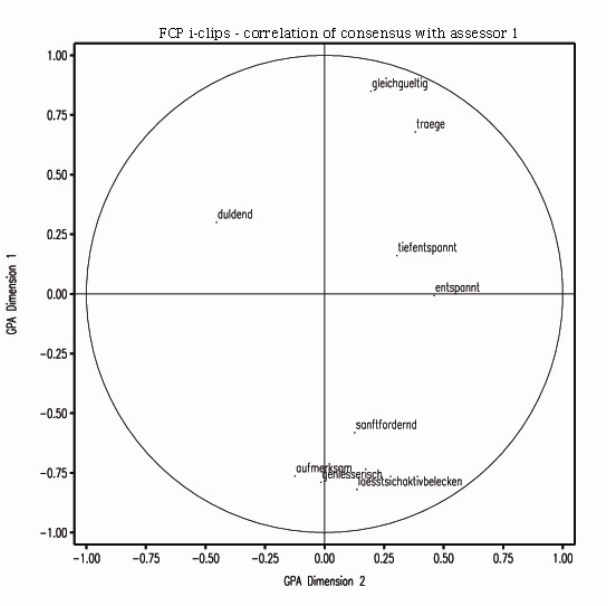

FCP h-clips - correlation of consensus with assessor 9
GPA Dimension 1
GPA Dimension 2
friedlich
ruhig
entspannt
geniessend
ungestoert
gleichgueltig
unwohl
unentschlossen
beengt
bedraengt
interessiert
neugierig
gereizt
zielstrebig
bestimmend
lebhaft

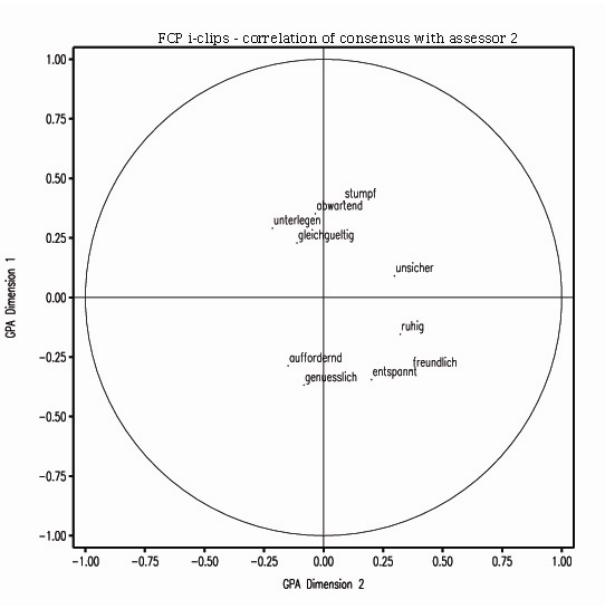

FCP h-clips - correlation of consensus with assessor 10
GPA Dimension 1
GPA Dimension 2
ruhig
entspannt
ratlos
fordernd
streitlustig
unruhig
aengstlich
angriffslustig
neugierig interessiert
unzufrieden
aggressiv
aufgeregt
aufmerksam
dominierend

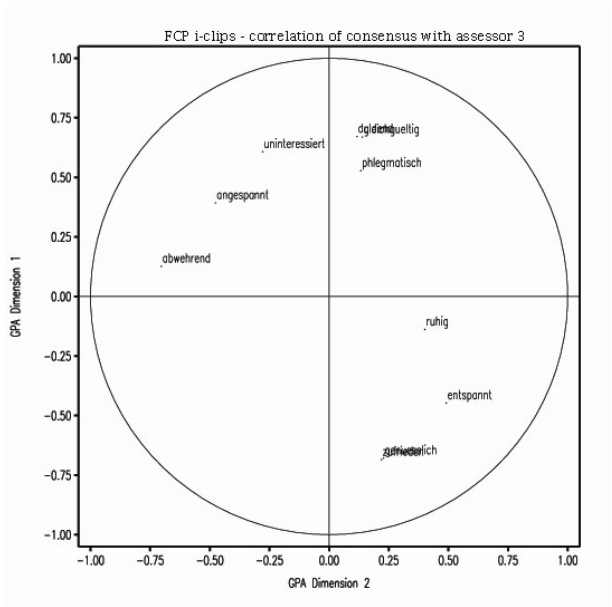

FCP h-clips - correlation of consensus with assessor 11
GPA Dimension 1
GPA Dimension 2
passiv
zufrieden
ruhig
entspannt
gelangweilt
aufetwaswartend
aufmerksam
angespannt
abgelenkt
unsicher
neugierig
genervt
gestresst
aktiv
unruhig
drohend
stuermisch
aggressiv

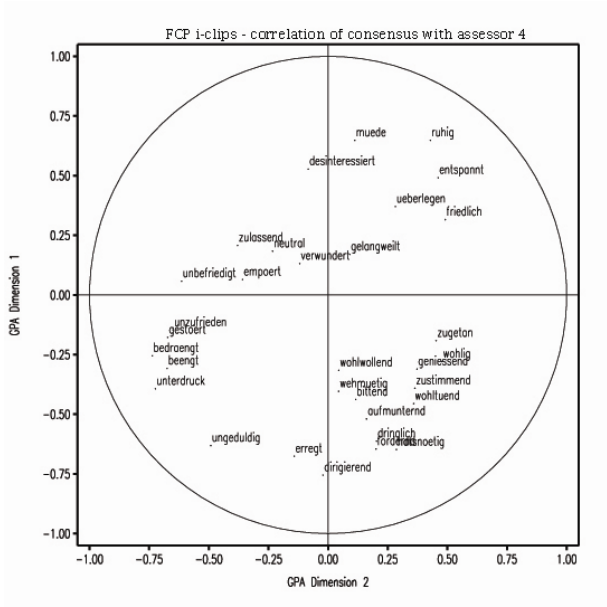

FCP h-clips - correlation of consensus with assessor 12
GPA Dimension 1
GPA Dimension 2
entspannt
geniessend
ruhig
vertrauensvoll
unentschlossen
aufmerksam
neugierig
aggressiv
frustriert
aktiviert
unsicher
unfreundlich
bestimmt
aktiv
angespannt
genervt
irritiert
unruhig

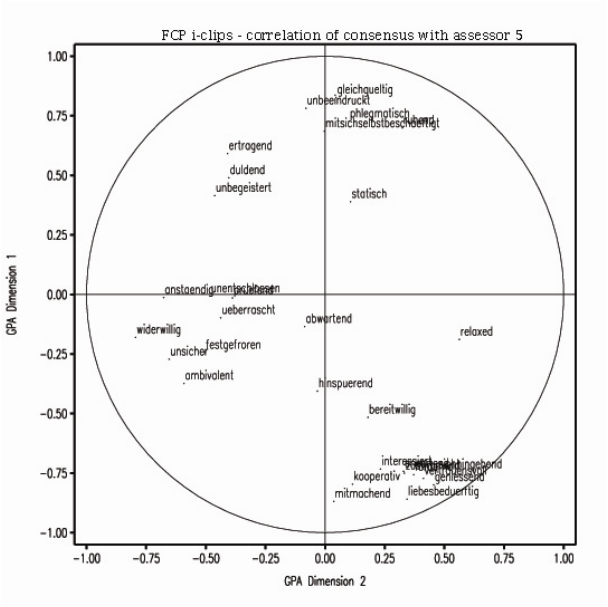
FCP Word Charts of 12 observers for i-clips

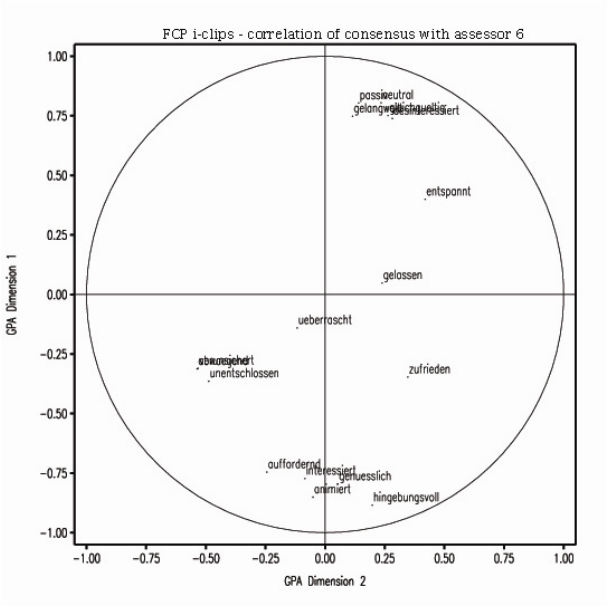


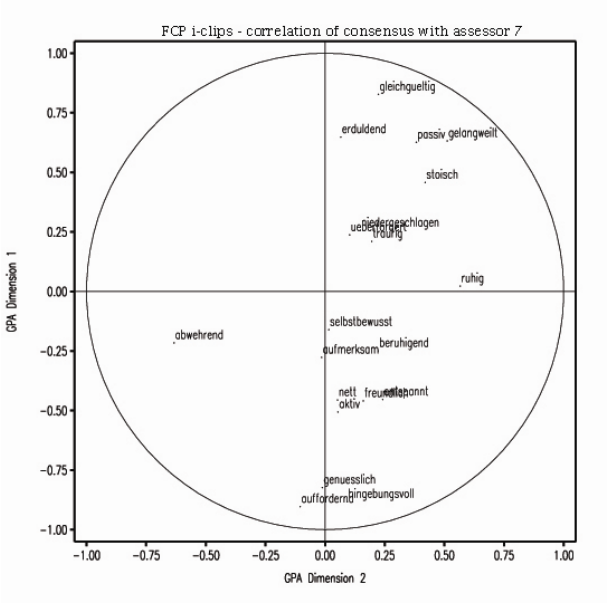


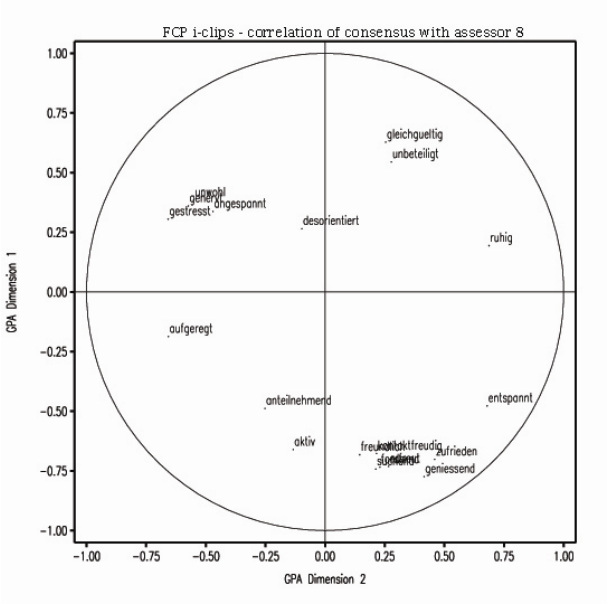


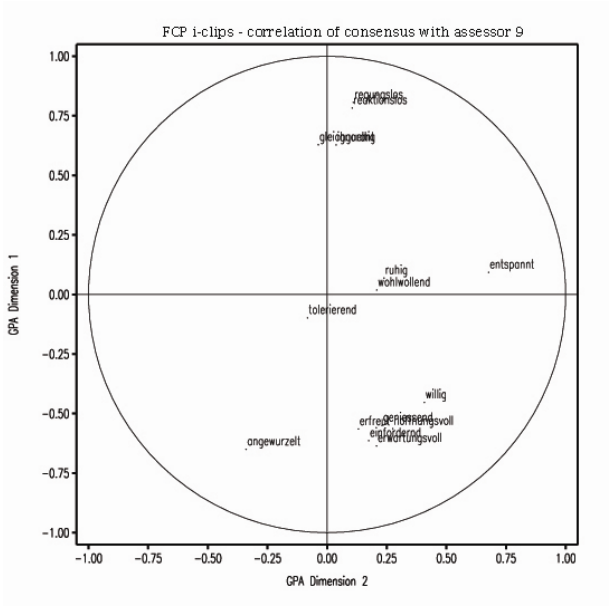


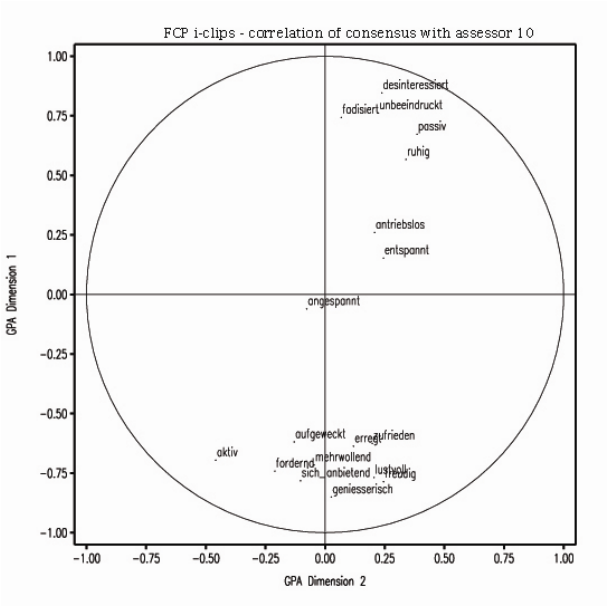

FCP i-clips - correlation of consensus with assessor 5
GPA Dimension 1
GPA Dimension 2
gleichgueltig
unbeeindruckt
phlegmatisch
mitsichselbstbeschaeftigt
ertragend
duldend
unbegeistert
statisch
ueberrascht
abwertend
widerwillig
relaxed
unsicher
festgefroren
ambivalent
hinspuerend
bereitwillig
kooperativ
mitmachend
liebesbeduerftig

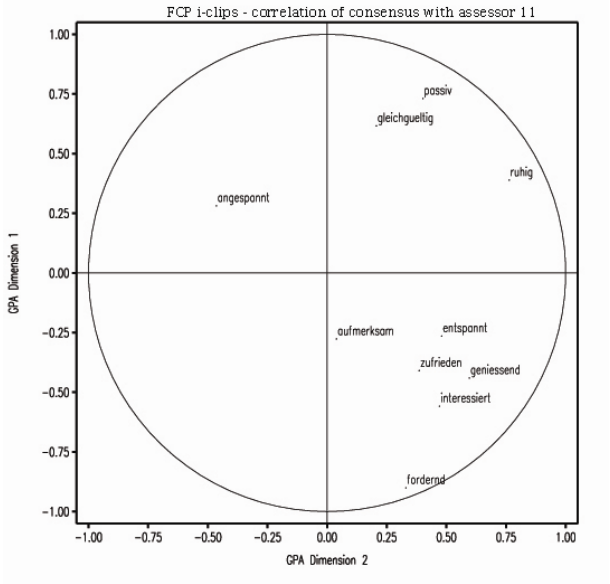

FCP i-clips - correlation of consensus with assessor 6
GPA Dimension 1
GPA Dimension 2
passiv
neutral
gelangweilt
gleichgueltig
desinteressiert
entspannt
gelassen
ueberrascht
unentschlossen
zufrieden
auffordernd
interessiert
genuesslich
animiert
hingebungsvoll

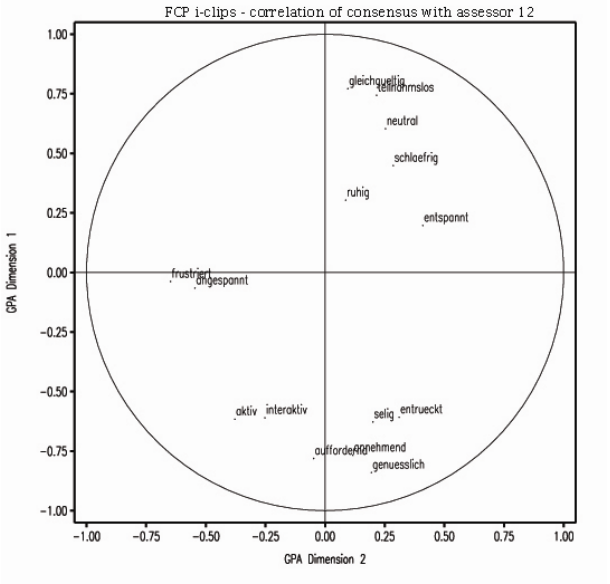

FCP i-clips - correlation of consensus with assessor 7
GPA Dimension 1
GPA Dimension 2
gleichgueltig
erduldend
passiv gelangweilt
stoisch
traurig
ruhig
selbstbewusst
abwehrend
beruhigend
aufmerksam
nett
oktiv
genuesslich
auffordernd
hingebungsvoll


FCP i-clips - correlation of consensus with assessor 8
GPA Dimension 1
GPA Dimension 2
gleichgueltig
unbeteiligt
gestresst
desorientiert
ruhig
aufgeregt
anteilnehmend
entspannt
aktiv
zufrieden
geniessend

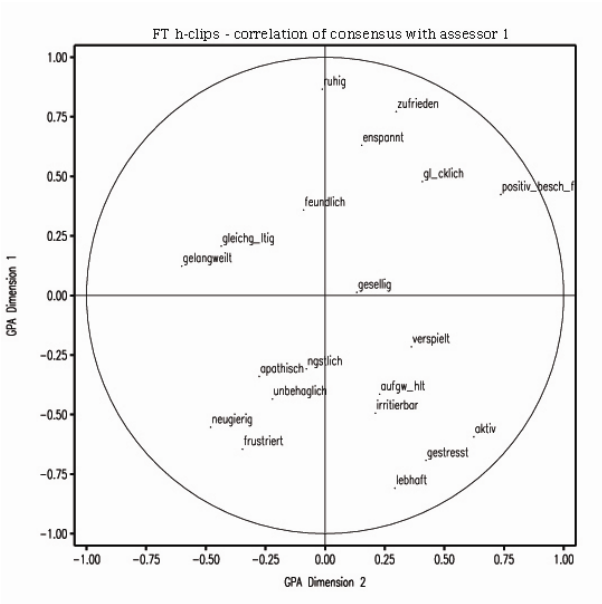

FCP i-clips - correlation of consensus with assessor 9
GPA Dimension 1
GPA Dimension 2
ruhig
wohlwollend
entspannt
tolerierend
willig
angewurzelt
erwartungsvoll

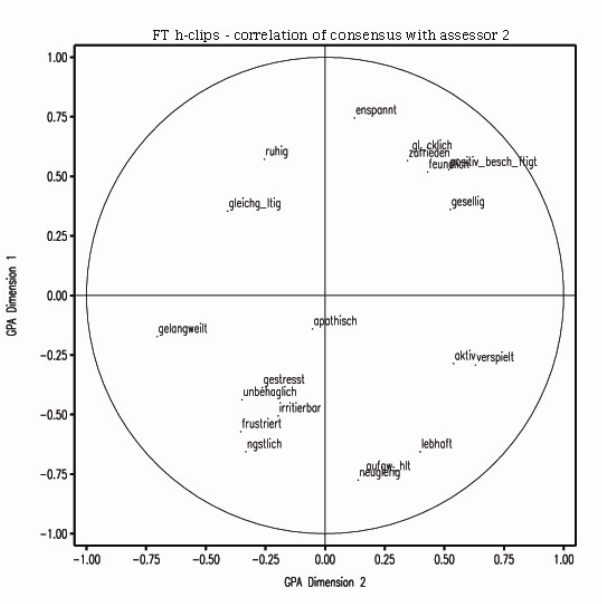

FCP i-clips - correlation of consensus with assessor 10
GPA Dimension 1
GPA Dimension 2
desinteressiert
fodisiert
unbeeindruckt
passiv
ruhig
antriebslos
entspannt
angespannt
aufgeweckt
zufrieden
aktiv
fordernd
mehrwollend
sich_anbietend
geniesserisch

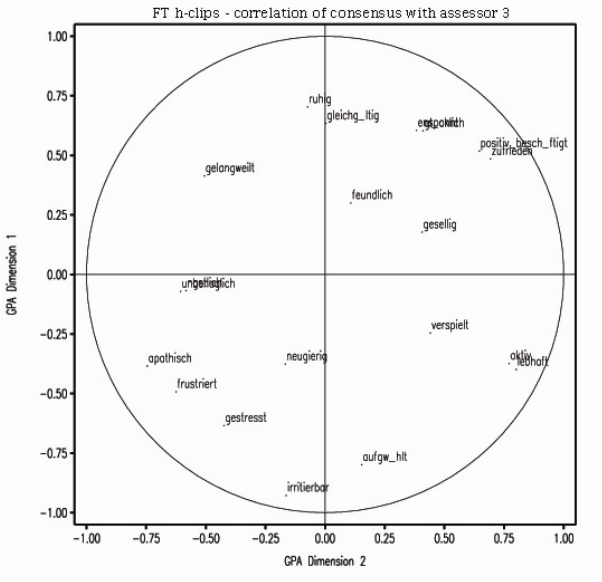

FCP i-clips - correlation of consensus with assessor 11
GPA Dimension 1
GPA Dimension 2
passiv
gleichgueltig
ruhig
angespannt
aufmerksam
entspannt
zufrieden
geniessend
interessiert
fordernd

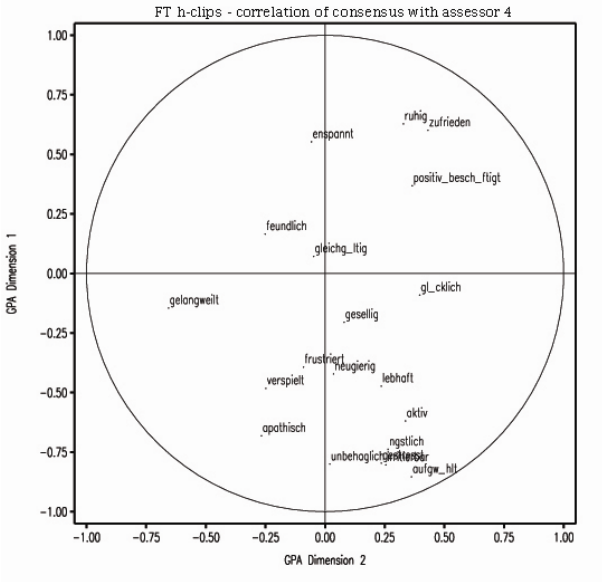

FCP i-clips - correlation of consensus with assessor 12
GPA Dimension 1
GPA Dimension 2
gleichgueltig
teilnahmslos
neutral
schlaefrig
ruhig
entspannt
frustriert
angespannt
aktiv
interaktiv
selig
entrueckt
aufforderend
angenehmend
genuesslich

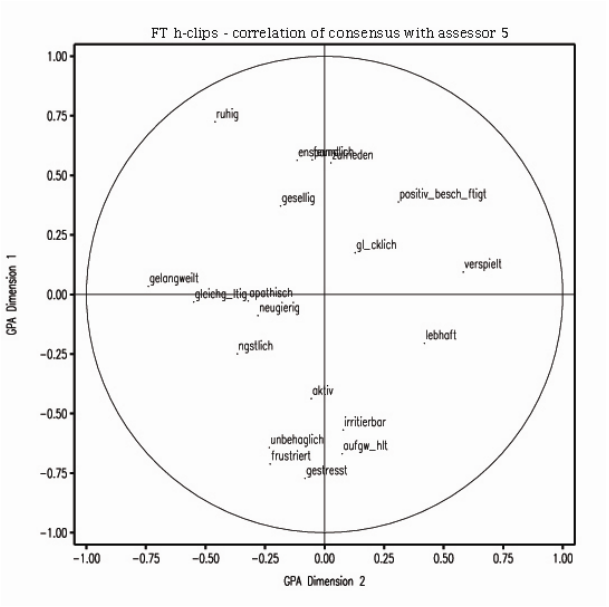
FT Word Charts of 12 observers for h-clips

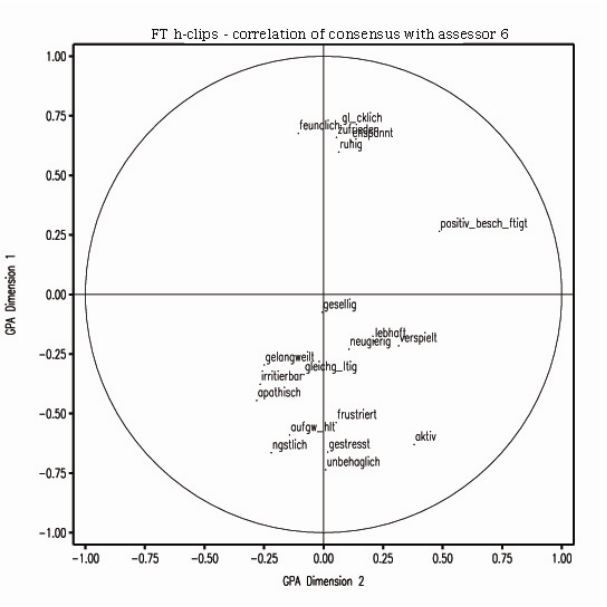


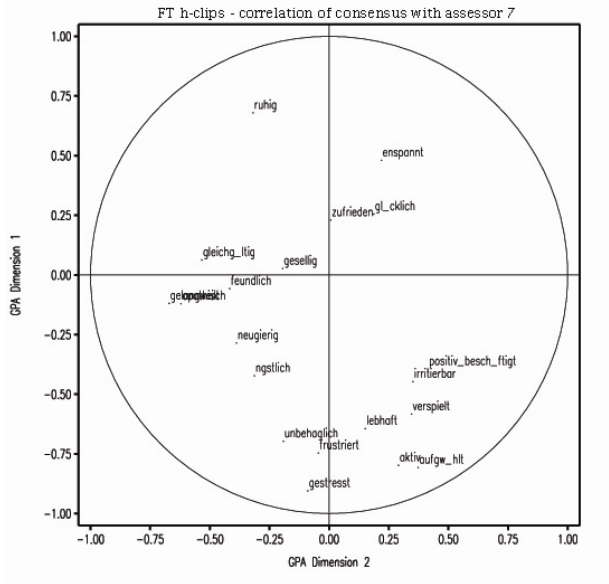


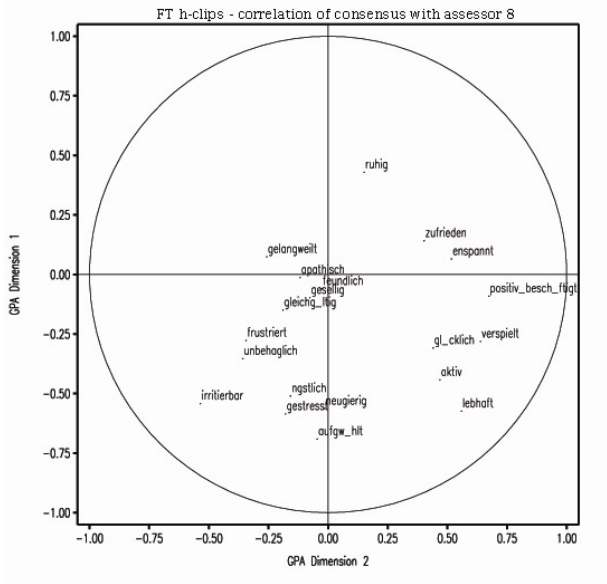


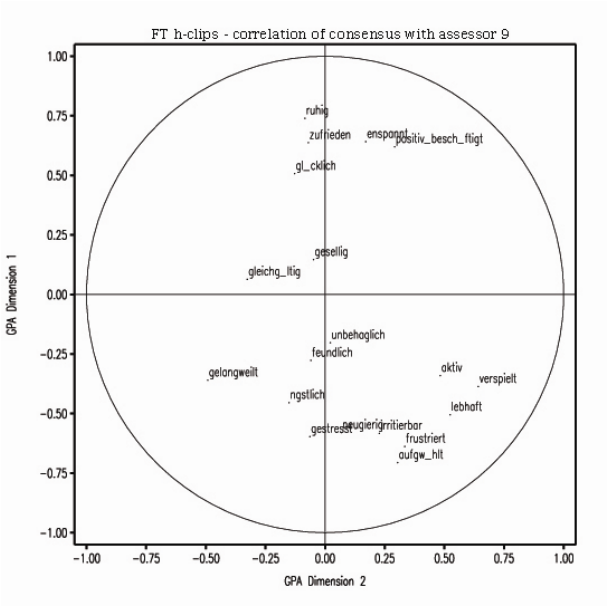


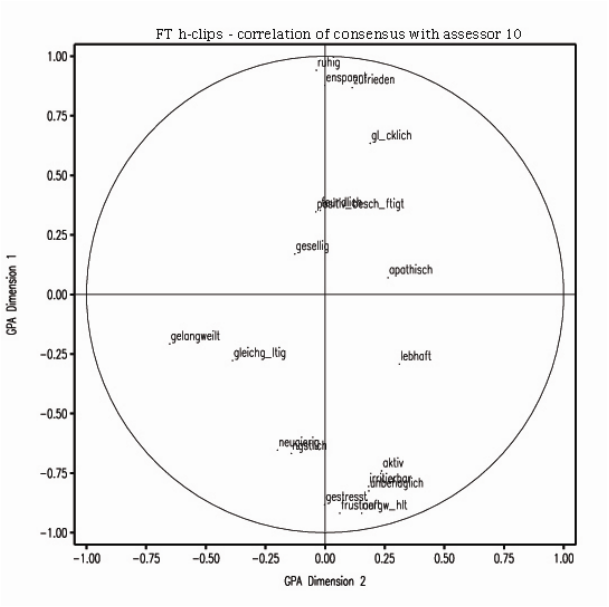

FT h-clips - correlation of consensus with assessor 5
GPA Dimension 1
GPA Dimension 2
ruhig
gesellig
positiv_besch_ftigt
gl_cklich
verspielt
gelangweilt
neugierig
lebhaft
aktiv
irritierbar
unbehaglich
frustriert
aufgw_hlt
gestresst

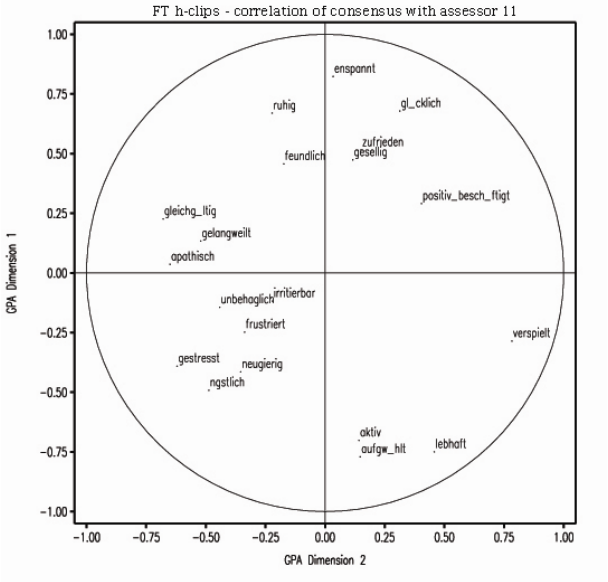

FT h-clips - correlation of consensus with assessor 6
GPA Dimension 1
GPA Dimension 2
ruhig
positiv_besch_ftigt
gesellig
lebhaft
verspielt
gelangweilt
irritierbar
apathisch
frustriert
aufgw_hlt
aktiv
gestresst
unbehaglich

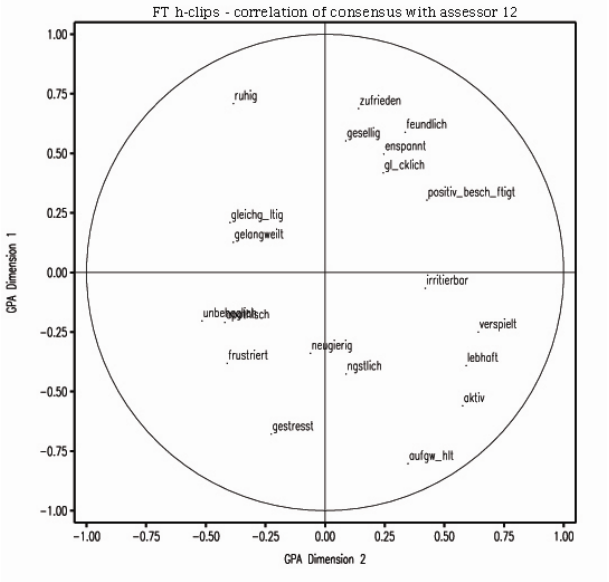

FT h-clips - correlation of consensus with assessor 7
GPA Dimension 1
GPA Dimension 2
ruhig
enspannt
zufrieden
gl_cklich
gesellig
feundlich
neugierig
positiv_besch_ftigt
irritierbar
verspielt
lebhaft
unbehaglich
frustriert
aufgw_hlt
gestresst


FT h-clips - correlation of consensus with assessor 8
GPA Dimension 1
GPA Dimension 2
ruhig
zufrieden
gelangweilt
enspannt
apathisch
feundlich
gesellig
positiv_besch_ftigt
gleichg_ltig
frustriert
gl_cklich
verspielt
unbehaglich
aktiv
irritierbar
ngstlich
gestresst
neugierig
lebhaft
aufgw_hlt

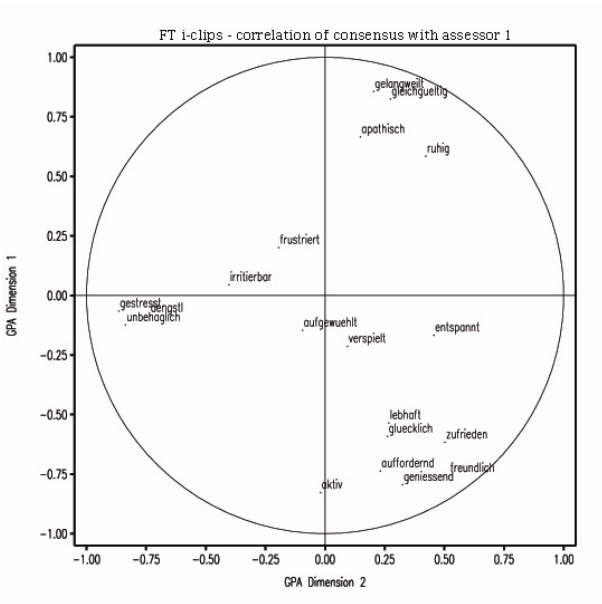

FT h-clips - correlation of consensus with assessor 9
GPA Dimension 1
GPA Dimension 2
ruhig
zufrieden
ensp
positiv_besch_ftigt
gl_cklich
gesellig
gleichg_ltig
unbehaglich
feundlich
gelangweilt
aktiv
verspielt
ngstlich
lebhaft
gestresst
neugierig
irritierbar
frustriert
aufgw_hlt

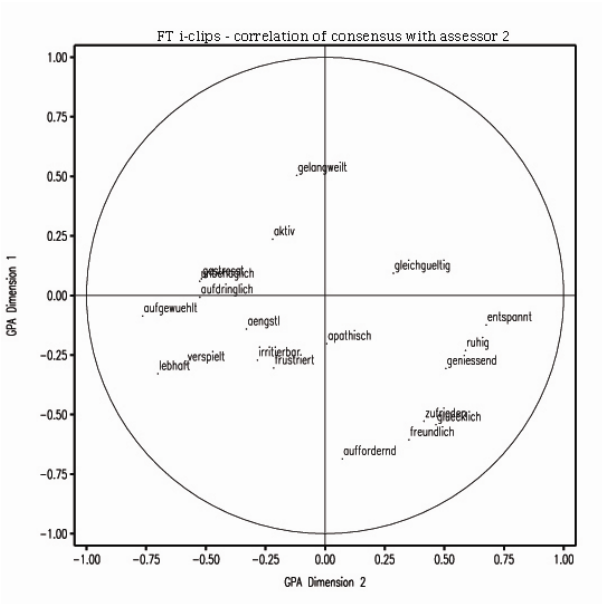

FT h-clips - correlation of consensus with assessor 10
GPA Dimension 1
GPA Dimension 2
ruhig
gl_cklich
gesellig
apathisch
gelangweilt
gleichg_ltig
lebhaft
aktiv
gestresst
aufgw_hlt

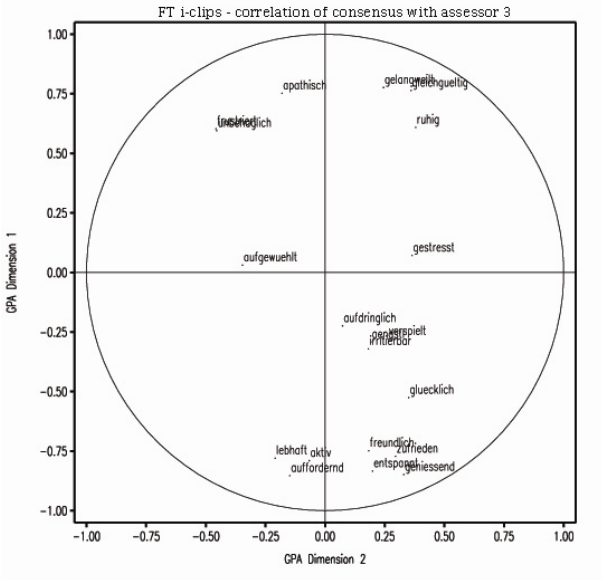

FT h-clips - correlation of consensus with assessor 11
GPA Dimension 1
GPA Dimension 2
enspannt
ruhig
gl_cklich
feundlich
zufrieden
gesellig
positiv_besch_ftigt
gleichg_ltig
gelangweilt
apathisch
irritierbar
unbehaglich
frustriert
verspielt
gestresst
neugierig
ngstlich
aktiv
aufgw_hlt
lebhaft

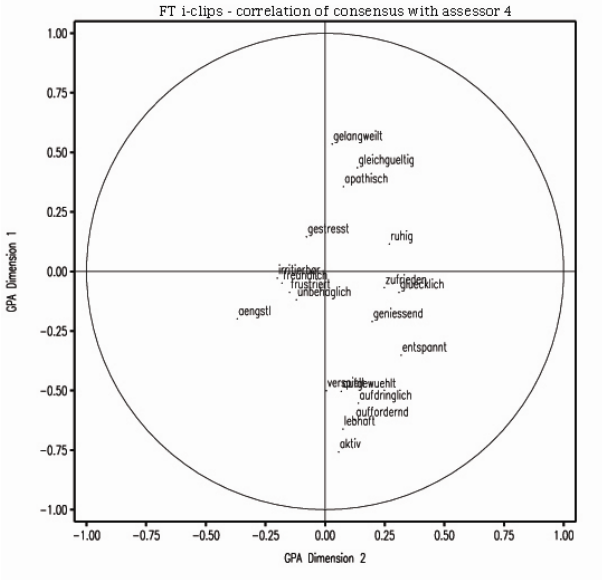

FT h-clips - correlation of consensus with assessor 12
GPA Dimension 1
GPA Dimension 2
ruhig
zufrieden
feundlich
gesellig
enspannt
gl_cklich
positiv_besch_ftigt
gleichg_ltig
gelangweilt
irritierbar
verspielt
frustriert
neugierig
ngstlich
lebhaft
aktiv
gestresst
aufgw_hlt

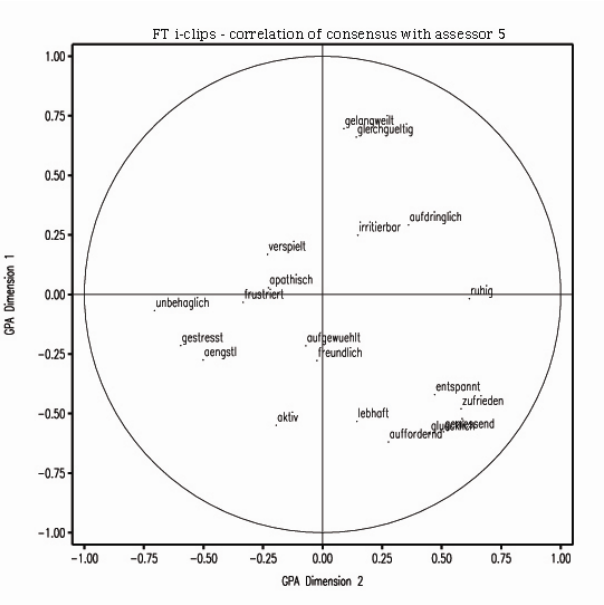
FT Word Charts of 12 observers for i-clips

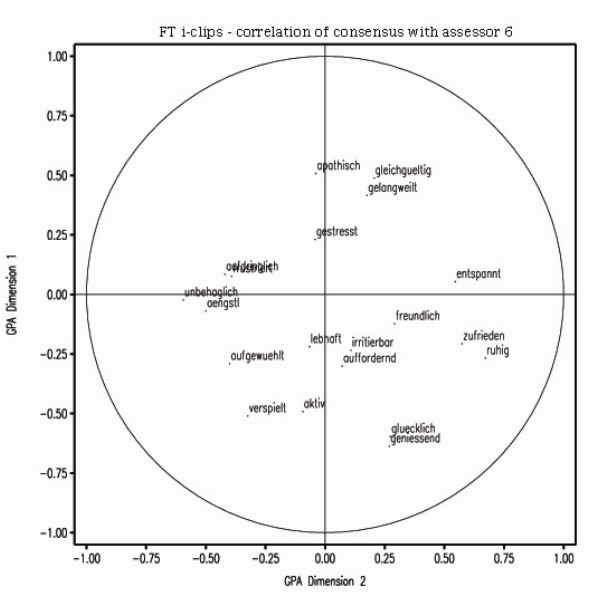


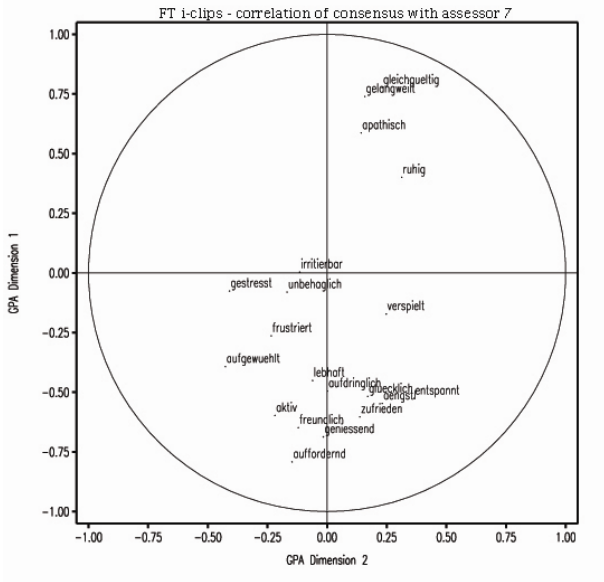


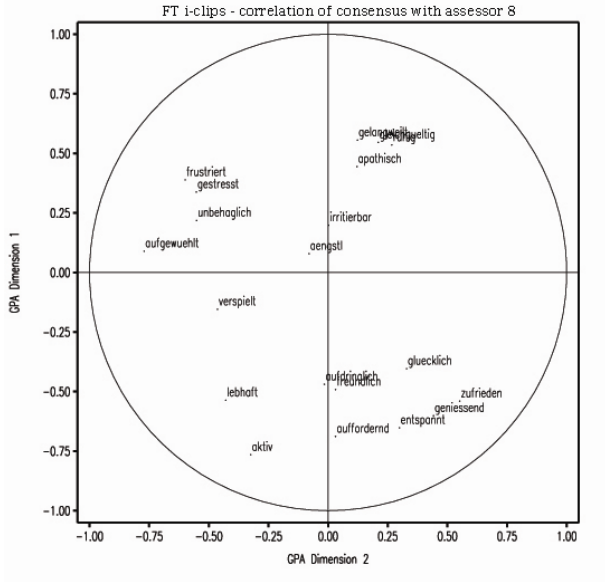


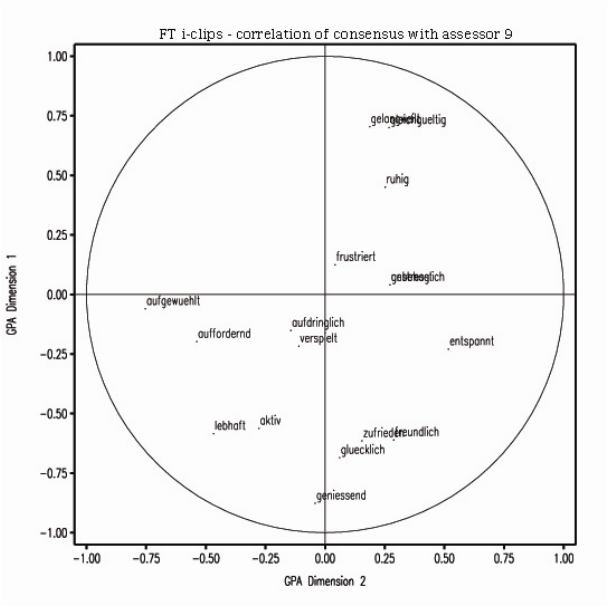


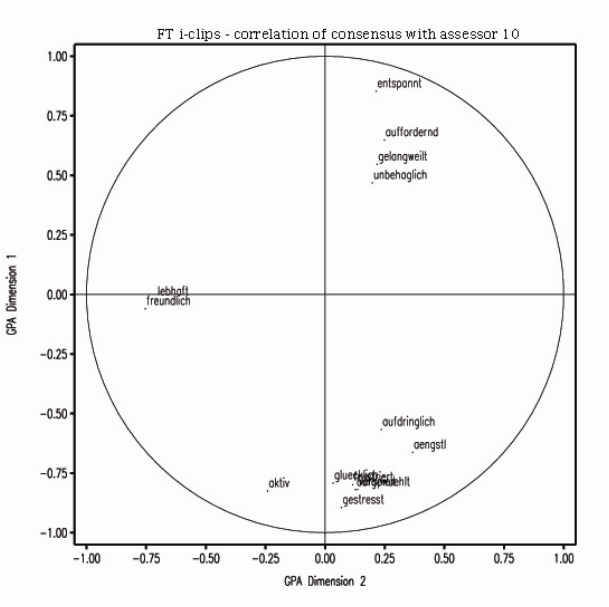

FT i-clips - correlation of consensus with assessor 5
GPA Dimension 1
GPA Dimension 2
gelangweilt
gleichgueltig
aufdringlich
irritierbar
verspielt
apathisch
frustriert
ruhig
unbehaglich
gestresst
aufgewuehlt
aengstl
freundlich
entspannt
zufrieden
lebhaft
aktiv
auffordernd

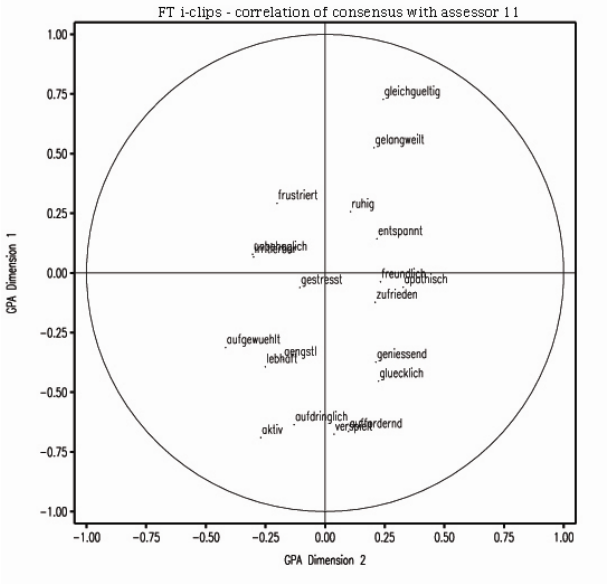

FT i-clips - correlation of consensus with assessor 6
GPA Dimension 1
GPA Dimension 2
apathisch
gleichgueltig
gelangweilt
gestresst
entspannt
unbehaglich
aengstl
freundlich
zufrieden
ruhig
lebhaft
irritierbar
aufgewuehlt
auffordernd
verspielt
aktiv
gluecklich
geniessend

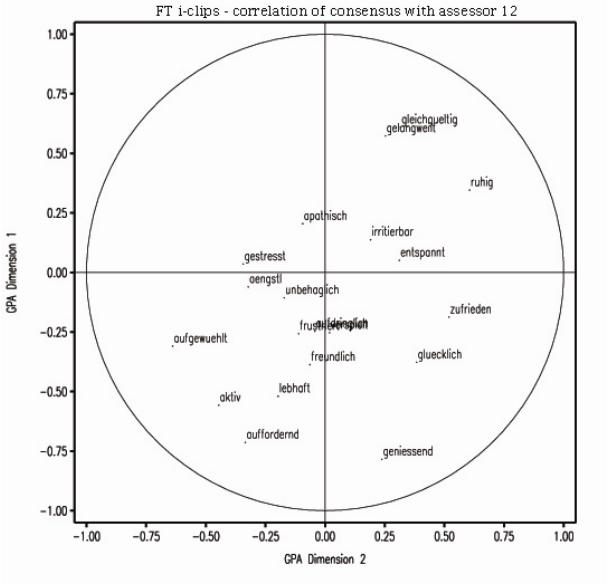

FT i-clips - correlation of consensus with assessor 7
GPA Dimension 1
GPA Dimension 2
gleichgueltig
gelangweilt
apathisch
ruhig
irritierbar
gestresst
unbehaglich
verspielt
frustriert
aufgewuehlt
lebhaft
aufdringlich
gluecklich
entspannt
aktiv
zufrieden
freundlich
geniessend
auffordernd


FT i-clips - correlation of consensus with assessor 8
GPA Dimension 1
GPA Dimension 2
apathisch
frustriert
gestresst
unbehaglich
irritierbar
aufgewuehlt
aengstl
verspielt
gluecklich
aufdringlich
freundlich
lebhaft
zufrieden
geniessend
entspannt
auffordernd
aktiv


FT i-clips - correlation of consensus with assessor 9
GPA Dimension 1
GPA Dimension 2
gleichgueltig
ruhig
frustriert
aufgewuehlt
aufdringlich
auffordernd
verspielt
entspannt
lebhaft
aktiv
freundlich
gluecklich
geniessend


FT i-clips - correlation of consensus with assessor 10
GPA Dimension 1
GPA Dimension 2
entspannt
auffordernd
gelangweilt
unbehaglich
lebhaft
freundlich
aufdringlich
aengstl
aktiv
gestresst


FT i-clips - correlation of consensus with assessor 11
GPA Dimension 1
GPA Dimension 2
gleichgueltig
gelangweilt
frustriert
ruhig
entspannt
gestresst
zufrieden
apathisch
aufgewuehlt
aengstl
lebhaft
geniessend
gluecklich
aufdringlich
aktiv


FT i-clips - correlation of consensus with assessor 12
GPA Dimension 1
GPA Dimension 2
gleichgueltig
ruhig
apathisch
irritierbar
gestresst
entspannt
aengstl
unbehaglich
zufrieden
aufgewuehlt
freundlich
gluecklich
lebhaft
aktiv
auffordernd
geniessend

9. References

Arnold, G.M.; Williams, A.A.; The use of generalised procrustes techniques in sensory analysis. Statistical Procedures in Food Research; Elsevier Applied Science, 1985, pp. 2233-2253

Azodanlou, R.; Darbellay, C.; Luisier, J.-L.; Villettaz, J.-C.; Amadò, R.; Development of a model for quality assessment of tomatoes and apricots. LWT Food Science and Technology , 2003, **36**, pp. 223-233

Berleant, D.; Zhang, J.; Using Pearson correlation to Improve envelopes around the distributions of functions. Reliable Computing, 2004, 10, pp. 139-161

Boissy, A.; Manteuffel, G.; Jensen, M.B.; Moe, R.O.; Spruijt, B.; Keeling, L.J.; Winckler, C.; Forkman, B.; Dimitrov, I.; Langbein, J.; Bakken, M.; Veissier, I.; Aubert, A.; Assessment of positive emotions in animals to improve their welfare. Physiology and Behavior, 2007, **92**, pp. 375-397

Botreau, R.; Bonde, M.; Butterworth, A.; Perny, P., Bracke, M.B.M; Capdeville, J.; Veissier, I.; Aggregation of measures to produce an overall assessment of animal welfare. Part 1: A review of existing methods. Anima, 2007, **1**, pp. 1179-1187

Broom, D. M.; Welfare Evaluation. Applied Animal Behaviour Science, 1997, **54**, pp. 21-23

Broom, D.M.; Animal Welfare: Concepts and Measurements. Journal of Animal Science, 1991, **69**, pp. 4167-4175

Broom, D.M.; Indicators of poor welfare. The British Veterinary Journal, 1986, **142**, pp. 524-526

Cook, N.B.; Mentink, R.L.; Bennett, T.B.; Burgi, K.; The effect of heat stress and lameness on time budget of lactating dairy cows. Journal of Dairy Science, 2007, 90, pp. 1674-1682

Dawkins, M.S.; Behaviour as a tool in the assessment of animal welfare. Zoology, 2003, **106**, pp. 383-387

Désiré, L.; Boissy, A.; Veissier, I.; Emotions in farm animals: A new approach to animal welfare in applied ethology. Behavioural Processes, 2002, **60**, pp. 165-180

Farm Animal Welfare Council; FAWC updates the five freedoms. 1992, The Veterinary Record pp. 131-357

Foley, J.P.; Judgment of facial expression of emotion in the chimpanzee. 1935, The Journal of Social Psychology, Vol 6, 31-67

Fraser, D.; Assessing animal welfare at the farm and group Level: The interplay of science and values. Animal Welfare, 2003, **12**, pp. 433-443

Gower, J.C.; Generalized Procrustes Analysis; Psychometrica, 1975, **40**, pp. 35-51

Grice, J.W.; Generalized Procrustes Analysis: A tool for exploring aggregates and persons. Applied multivariate research, **13**, 2009, pp. 93-112

Kuhar, C.W; Stoinski, T.S.; Lukas, K.E.; Maple, T.L.; Gorilla Behaviour Index revisited: Age, housing and behaviour. Applied Animal Behaviour Science, 2006, **96**, pp. 315-326

Lund, V.; Ethics and animal welfare in organic animal husbandry – an interdisciplinary approach. Doctor's Dissertation, SLU Uppsala; 2002

McDonnell, S.M.; Kristula, M.A.; No effect of drinking water temperature (ambient vs. chilled) on consumption of water during hot summer weather in ponies. Applied Animal Behaviour Science, 1996, **49**, pp. 159-163

Meagher, R.K.; Validity and values as a tool for animal welfare research, Applied Animal Behaviour Science, 2009, **119**, pp. 1-14

Minero, M.; Tosi, M.V.; Canali, E.; Wemelsfelder, F.; Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. Applied Animal Behaviour Science, 2009, **116**, pp. 74-81

Napolitano, F.; De Rosa, G.; Braghieri, A.; Grasso, F.; Bordi, A.; Wemelsfelder, F.; The qualitative assessment of responsiveness to environmental challenge in horses and ponies. Applied Animal Behaviour Science, 2008, **109**, pp. 342-354

Neisen, G.; Wechsler, B.; Gygax, L.; Choice of scan-sampling intervals - an example with quantifying neighbours in dairy cows. Applied Animal Behaviour Science, 2009, **31**, pp. 143-140

O'Driscoll, K.; Boyle, L.; Hanlon, A.; A brief note on the validation of a system for recording lying behaviour in dairy cows. Applied Animal Behaviour Science, **111**, 2008, pp. 195–200

Qannari, E.M.; Wakeling, I.; MacFie, H.J.H.; A hierarchy of models for analysing sensory Data. Food Quality and Preferences, 1995, **6**, pp. 309-314

Rousing, T.; Wemelsfelder, F.; Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. Applied Animal Behaviour Science, 2006, **101**, pp. 40-53

Sinesio, F.; Moneta, E.; Sensory evaluation of walnut fruit. Food Quality and Preference, 1997, **8**, No. 1, pp. 35-43

Thamke, I.; Dürrschmid, K.; Rohm, H.; Sensory description of dark chocolates by consumers. LWT – Food Science and Technology, 2009, **42**, pp. 534-539

Villarino, B.J.; Dy, L.M.; Lizada, C.C.; Descriptive sensory evaluation of virgin coconut oil and refined, bleached and deodorized coconut oil. LWT – Food Science and Technology, 2004, **40**, pp. 193-199

Watanabe, S.; Animal psychology contributes to animal welfare. Applied Animal Behaviour Science, 2007, **106**, pp. 193-202

Wemelsfelder, F.; The scientific validity of subjective concepts in models of animal welfare. Applied Animal Behaviour Science, 1997, **53**, pp. 75-88

Wemelsfelder, F.; Hunter, E.A.; Mendl, M.T.; Lawrence, A.B.; The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. Applied Animal Behaviour Science, 2000, **67**, pp. 193-215

Wemelsfelder, F.; Hunter, T.E.A.; Mendl, M.T.; Lawrence A.B.; Assessing the "whole animal": A Free Choice Profiling Approach. Animal Behaviour, 2001a, **62**, pp. 209-220

Wemelsfelder, F.; Lawrence, B.A.; Qualitative assessment of animal behaviour as an on-farm welfare-monitoring tool. Animal Science, 2001b, **30**, pp. 21-25

Wemelsfelder, F.; Knierim, H.; Westerath, H.S.; Lentfer, T.; Staack, M.; Sandilands, V.; Assessment of animal welfare measures for layers and broilers. Welfare Quality Reports No. 9, edited by B. Forkman and L. Keeling; 2009a, pp. 113-119

Wemelsfelder, F.; Millard, F.; Assessment of animal welfare measures for sows, piglets and fattening Pigs. Welfare Quality Reports No. 10, edited by B. Forkman and L. Keeling; 2009b, pp. 213-219

Wemelsfelder, F.; Millard, F.; De Rosa, G.; Napolitano, F.; Assessment of animal welfare measures for dairy cattle, beef bulls and veal calves. Welfare Quality Reports No. 11; 2009c, edited by B. Forkman and L. Keeling; pp. 215-224

Wemelsfelder F.; Nevison, I.; Lawrence, A.B.; The effect of perceived environmental background on qualitative assessments of pig behaviour. Animal Behaviour, 2009d, **78**, pp. 477-484

Wielebnowski, N.; Behavioral differences as predictors of breeding status in captive cheetahs. Zoo Biology, 1999, **18**, pp. 335-349

Wilkinson, C.; Schipper, M.; Leguijt, T.; Weighted analysis for missing values in generalized procrustes analysis. Food Quality and Preferences, 2000, **11**, pp. 85-90

Williams, A.A.; Langron, S.P.; The use of free-choice profiling for the evaluation of commercial ports. Journal of the Science of Food and Agriculture, 1984, **35**, pp. 558-568

Wu, W.; Guo, Q.; De Jong, S.; Massart, D.L.; Randomisation test for the number of dimensions of the group average space in generalised procrustes analysis. Food Quality and Preference, 2003, **13**, pp. 191-200

Xiong, R.; Blot, K.; Meullenet, J.F; Dessirier, J.M.; Permutation tests for generalized procrustes analysis. Food Quality and Preference, 2008, **19**, 146-155