Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Evaluation of assessment method for the trait Spirit in breeding field tests for Icelandic horses

Heiðrún Sigurðardóttir

Examensarbete / Swedish University of Agricultural Sciences
Department of Animal Breeding and Genetics

496

Uppsala 2016

Master's Thesis, 60 hp

Master's Programme
– Animal Science

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Evaluation of assessment method for the trait Spirit in breeding field tests for Icelandic horses

Heiðrún Sigurðardóttir

Supervisor:
Susanne Eriksson, SLU, Department of Animal Breeding and Genetics
Elsa Albertsdóttir, The Farmers Association of Iceland, Reykjavik

Examiner:
Åsa Viklund, SLU, Department of Animal Breeding and Genetics

Credits: 60 hp
Course title: Degree project in Animal Science
Course code: EX0713
Programme: Master's programme - Animal Science
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2016
Name of series, no: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 496
On-line publication: http://stud.epsilon.slu.se

Keywords: Icelandic horse, temperament, field test, questionnaire, survey, culling, genetic analysis, heritability

Abstract

The temperament trait *Spirit* is included in the official breeding goal for Icelandic horses. Spirit is subjectively assessed at breeding field tests, and is included in the genetic evaluation for Icelandic horses. The assessment method for spirit has been discussed within the Icelandic horse society for several years where some consider it to favour nervous and tense horses instead of cooperative and stable horses. In 2014, the Horse Breeders Association in Iceland initiated a project for a trial period with additional assessments pertaining to temperamental suppleness by breeding judges at breeding field tests, as means to improve the assessment of spirit.

The aim of this thesis was to evaluate the current assessment method for spirit. Genetic parameters were estimated for all temperament traits assessed at breeding field tests as well as general temperament traits assessed at home by owners. The consistency between assessments made by breeding judges and riders at breeding field tests was investigated, as well as the consistency between assessments made by owners at home and assessments made by breeding judges at breeding field tests. The frequency of culling due to temperament faults was also investigated by gathering information from owners of culled horses.

The estimated heritabilities of the temperament traits assessed at breeding field tests ranged from 0.00 to 0.76, and differed depending on whether riders or judges assessed the trait. It seemed that the riders were better suited to assess some traits. It was concluded that the new suppleness traits assessed at breeding field tests provide additional information supporting the assessment of spirit but trait definitions need to be improved. Estimated heritabilities for general temperament traits ranged from 0.00 to 0.56. The highest heritability was estimated for training level which describes the genetic potential of the horse to respond to training. Genetic correlations between the general temperament traits and spirit were estimated on a wide range (0.01-0.97) where some traits pertaining to general nerve strength and training response were negatively correlated to spirit. It was concluded that the score for spirit describes only part of the general temperament of the horse. Nevertheless the majority of horse owners in this study were satisfied with the assessment method for spirit. Approximately one third of culled horses were culled at least partly due to temperament faults whereof the majority were young horses that were withdrawn from further training. These results may indicate a strong preselection based on temperament during the process of breaking in young horses.

Ágrip

Hið opinbera ræktunarmarkmið íslenska hestsins felur meðal annars í sér eiginleikann *Vilji og geðslag*. Eiginleikinn er metinn á huglægan hátt af kynbótadómurum á kynbótasýningum og er einn af þeim eiginleikum sem kynbótamatið er byggt á. Matsaðferð eiginleikans hefur í gegnum tíðina verið umdeild meðal hestamanna, en margir líta svo á að aðferðin styðji ekki nógu vel við þjál og meðfærilega hross fyrir hinn almenna reiðmann. Í framhaldi af umræðunni árið 2014, kom Fagráð í hrossarækt á fót tímabundnu verkefni sem fól í sér mat á þjálnieiginleikum hrossa. Þjálnimatið var framkvæmt af kynbótadómurum samhliða mati á reiðhæfileikum í kynbótadómi og var markmið þess að reyna að bæta matið á vilja og geðslagi.

Markmið þessarar rannsóknar var að meta núverandi matsaðferð vilja og geðslags. Erfðastuðlar voru reiknaðir fyrir vilja og geðslag og þjálnieiginleika metna af bæði dómurum og knöpum á kynbótasýningum, og fyrir almenna geðslagseiginleika metna af eigendum kynbótahrossa í daglegu umhverfi þeirra. Samræmi milli mats dómara og knapa var kannað, ásamt samræmi milli mats eigenda og þess mats sem sömu hross hlutu fyrir vilja og geðslag í kynbótadómi. Tíðni slátrunar vegna geðslagsbresta var einnig könnuð.

Mat arfgengis vilja og geðslags og þjálnieiginleika var á bilinu 0.00 til 0.76 og var nokkuð breytilegt eftir því hvort eiginleikarnir voru metnir af dómurum eða knöpum, en svo virtist sem knapar ættu auðveldara með að meta suma eiginleika. Ályktað var að þjálnimatið veitir viðbótarupplýsingar sem styrkja matið á vilja og geðslagi en betri skilgreininga þjálnieiginleikanna er þörf. Mat arfgengis almennra geðslagseiginleika var á bilinu 0.00 til 0.56. Hæst metna arfgengið reyndist vera fyrir eiginleikann þjálfunarstig sem lýsir líffræðilegum möguleikum hrossins að svara þjálfun. Erfðafylgni milli vilja og geðslags og almennra geðslagseiginleika var metin á víðu bili (0.01-0.97) þar sem fylgni nokkurra eiginleika sem lutu að almennum taugastyrk og svörun við þjálfun var neikvæð við vilja og geðslag. Ályktað var að einkunn vilja og geðslags lýsir aðeins hluta af hinu almenna geðslagi hrossins. Engu að síður var meirihluti eigenda kynbótahrossa sem tóku þátt í þessari könnun ánægðir með matið á vilja og geðslagi. Um þriðjungur þeirra hrossa sem slátrað var hér á landi á ákveðnu tímabili, var slátrað vegna geðslagsvandamála. Meirihluti þessa hóps voru ung hross sem hætt var með í tamningu eða þjálfun. Þessar niðurstöður gefa til kynna að töluvert forval byggt á geðslagi fari fram í frumtamningarferli ungra hrossa.

Table of Contents

List of Tables

List of Figures

1 Introduction

Well defined and adequately administrated breeding programs for temperament traits in horses is imperative given the relevance to human safety, economics, management and animal welfare. Assessments for temperament traits are commonly carried out in horse breeding programs but only few breeding associations use the obtained information to estimate breeding values for temperament traits to use as guidelines in selection (von Borstel, 2013). This can partly be explained by the lack of definitions in existing guidelines for temperamental traits, but many attempts have been made by breeding authorities to improve assessments procedures and develop more objective assessments of temperament traits (von Borstel, 2013). Most studies conducted in relevance to this development indicate that temperament traits are rated by horse enthusiasts as more important than any of the conventional performance traits (von Borstel, 2013).

Icelandic horses are known for their calm temperament as well as their ability to perform in five different gaits. The main focus in breeding is consequently on well-tempered five-gaited horses with good, functional conformation suited for leisure riding and competition (FEIF, 2016a). The official breeding goal for Icelandic horses emphasises breeding of energetic, fiery horses with attractive movements and the ability to perform in great speed in most gaits. It contains 15 traits of conformation and riding ability, including the temperament trait *Spirit*. Assessment for spirit is carried out at standardised breeding field tests simultaneously with riding ability traits, where the whole performance is taken into account, and is subjectively assessed according to a defined judging scale. Spirit has considerable weight in the selection criteria and is included in the breeding evaluation for Icelandic horses. The assessment procedure for spirit has been a topic of discussion within the Icelandic horse society for several years. The focus of these discussions has been on whether the current assessment method for spirit actually supports breeding of well-tempered riding horses suited for the general rider, or if it supports breeding of tense horses only suited for a small group of professional riders. In 2014, the official Horse Breeders Association council in Iceland initiated a project involving additional assessments made by breeding judges at breeding field tests pertaining to temperamental suppleness of breeding horses, as means to improve the assessment of spirit.

The focus of this thesis was to evaluate the current assessment method for spirit, including the assessment of temperamental suppleness, by looking into consistency between assessments made by breeding judges, riders and owners, and to estimate the frequency of culling due to temperament faults in order to gain information on temperament in horses that are usually not included in the preselected group of breeding horses that are presented at breeding field tests.

2 Literature review

2.1 The Icelandic horse

Historical records state that since the time of settlement in Iceland in the late 8th century, the horse has been an important part of the Icelandic history and culture. Since the time of settlement the horses have remained isolated in Iceland and in the late 18th century regulations were established prohibiting importation of foreign genetic material (Björnsson & Sveinsson, 2004). Therefore the Icelandic horse is the only horse breed in Iceland and is considered pure-bred (Adalsteinsson, 1981).

Organised breeding in Iceland arose in the late 18th – early 19th century. In the beginning the emphasis was on two separate breeding goals which aimed at breeding draught horses and riding horses. Organised breeding with two breeding goals soon came to an end and as mechanisation of agriculture and transport became widespread in Iceland around 1950, the goal of breeding riding horses exclusively was established. The goal was to breed five-gaited horses, with good conformation and temperament, suited for riders of all kinds (Arnórsson, 2004). A scoring system for evaluation of breeding horses was established and consisted of scoring of separate traits, including several conformation and riding ability traits, which was considered unique at that time (Arnórsson, 2004). The essence of the breeding goal and the structure of the scoring system have remained similar since then but with more scientific research and knowledge gained over time, several adjustments have been made.

2.1.1 Horse keeping in Iceland

Horse keeping in Iceland is in many ways unique compared to other countries. In 2015, there were around 98,000 live horses registered in Iceland and about 11,500 people registered within the horse associations in Iceland (FEIF, 2016b) which suggests that each person within the horse associations owned around 8-9 horses each on average. About 7000 foals are born in Iceland annually and are sired by approximately 1000 stallions (Sigurðardóttir, 2012). Many of these horses inhabit and graze the extensive pastures of Iceland all year round and are facilitated with additional feed during the coldest months. This applies to most of the breeding mares, young horses and older riding horses. Younger breeding horses and riding horses are commonly kept in stables during the winter while being trained and are released into the pastures during summer or autumn. Horse keeping in Iceland could generally be described as non-profit, as it is expensive to keep horses and conventional competitions do not involve large amount of prize money or earnings. Profitable horse breeding in Iceland depends on marketing and sales and selection of horses can be quite rigorous. Procedures such as castration and culling of young horses of lower quality are commonly applied.

Horse owners and breeders in Iceland breed horses for various reasons. The most numerous groups of breeders are those who breed horses in small amount for their own use, and the so called ‘elite’ breeders who breed horses for the purpose of selling them. Other groups of breeders include those who breed horses for meat production, often combined with collection of blood hormones from the broodmares and those who breed horses used for guided riding tours. Breeding and competition horses are the most valuable horses and are most easily sold, both within the country as well as abroad. Therefore, it can be assumed that the ‘elite’ breeders are more actively involved in the organised breeding system performing selection based on strict criteria, as good pedigree and breeding assessments is valuable information as well as competition records when selling horses. Horse owners and breeders of both groups often utilise the service provided by privately operated horse training stations for braking in young horses around the age of 4 years, especially the more valuable and promising horses. These training stations are usually operated by professional riders employing younger riders and trainees as co-trainers. Promising horses for future breeding or competition career are often subjected to further training at these training stations. Other horses return to their owners who either continue with the training to acquire a good riding horse or decide to cull the horse, especially if it has displayed some unwanted behaviour or lack of riding ability. The selection of young horses for further training, either in these training stations or by their owners, is usually based on economic reasons.

It is common procedure in Iceland to castrate young colts around the age of 10 months, especially within the group of horse owners who breed horses for the own use. The ‘elite’ breeders are not as likely to castrate the colts at such young age, as the value of geldings is less than of stallions. These young colts and stallions are often subjected to unofficial young stallion assessments performed by breeding judges, and decisions on whether to castrate or not are based on the outcome as well as pedigree data. In some cases the young stallions are castrated after being broken in, especially if their performance does not support their potential for breeding.

Icelandic horses are considered to be long lasting and having high median length of life, especially with the high frequency of voluntary culling taken into account. The median length of life has been reported approximately 18 years for mares and 17 years for stallions or geldings, who had a much lower survival in the analysed sample (Sigfússon, 2003). The oldest Icelandic horse was reported to have lived until the age of 57 years (Björnsson & Sveinsson, 2004), but commonly old horses are culled when they have reached the point where they can no longer serve their purpose, usually well before the age of 30 years. Previous studies on culling rate and longevity in Icelandic horses reported that the most frequent reasons for culling are musculoskeletal system disorders (12%-60%), lack if genetic potential (20%), temperamental disorders (16%), accidents (12%) and bad performance (7%-11%) (Björnsdóttir, Árnason & Lord, 2003; Sigfússon, 2003).

2.1.2 Breeding goal

The official breeding goal for Icelandic horses describes a healthy, fertile and durable horse – a robust horse with great variation of coat colours within the breed and a preferred size of 135 – 145 cm measured at withers with a rod. The breeding goal for conformation in general aims at breeding light-bodied horses with emphasis on strength, flexibility and a muscular body which facilitates excellent gait performance and a naturally good head carriage. The breeding goal for riding abilities in general aims at breeding a versatile, consistent in gaits and reliable horse with good, clear gaits and an excellent, lively temperament (FEIF, 2016a).

Table 1. Conformation and riding ability traits included in the selection criteria and the weight assigned to each trait

Conformation traits	Weight	Riding ability traits	Weight
Head	3%	Tölt	15%
Neck, withers and shoulders	10%	Trot	7.5%
Back and croup	3%	Pace	10%
Proportions	7.5%	Gallop	4.5%
Legs (quality)	6%	Spirit	9%
Legs (joints)	3%	Form under rider	10%
Hooves	6%	Walk	4%
Mane and tail	1.5%		
Total	*40%*	*Total*	*60%*

The breeding goal is more closely defined as the traits of the selection criteria. The selection criteria includes 15 traits which are subjectively scored at breeding field tests according to a linear judging scale, where the highest score describes the breeding objective for each trait. Included in the selection criteria are eight conformation traits and seven riding ability traits, each assigned specific weight summarizing up to 40% weight for conformation traits of the total score and 60% weight for riding ability traits. The traits and the weight assigned to each one are presented in Table 1. The judging scale and the breeding goal for each trait are described in more detail in *FEIF Breeding Rules and Regulations* (FEIF, 2016a).

2.1.3 Breeding system

The structure of the breeding system for Icelandic horses consists of definition of breeding objectives, registration of horses, assessments of horses performed at breeding field tests, estimation of breeding values and selection performed by the breeders.

The Farmers Association of Iceland (FAI) is the head supervisor for the organised horse breeding system in Iceland. The official Horse Breeders Association council (*i. Fagráð*) works within the FAI and forms the structure of breeding and development in the breeding system in Iceland, defines the breeding objective and sets rules regarding breeding evaluations according to Icelandic Agriculture laws (no. 70/1998). The chairman of the Horse Breeders Association is responsible for guidance and administering the breeding system in Iceland. The International Federation of Icelandic Horse Associations (FEIF) represents Icelandic horse associations in 20 countries worldwide. Its goal is to promote the Icelandic horse, provide guidelines and unify standards regarding breeding, competition, education and other aspects of horsemanship of the Icelandic horse (FEIF, 2016c). WorldFengur (www.worldfengur.com), the studbook of origin for the Icelandic horse is published on the worldwide web and hosts information on Icelandic horses in all membership countries of FEIF. WorldFengur is administrated by the FAI and FEIF and includes information of pedigree, offspring, assessments, estimated breeding values, owners, breeders, competition data and more for around 388.000 Icelandic horses worldwide (WorldFengur, 2016).

Breeding values for the Icelandic horses are estimated using multitrait BLUP animal model, initiated by Þorvaldur Árnason in 1983 (Árnason, 1984). Since then the breeding evaluation has evolved and expanded to include breeding assessments made in 12 membership countries of FEIF (Kristjánsson, 2014). The evaluation today is based on available information for each individual, comprising individual assessments made at breeding field tests and assessments of related individuals, including progenies. The breeding evaluation is calculated annually in the autumn, with additional calculation in the summer in relation to the biannual National Horse Show of Iceland. The breeding evaluation adjusts for effects of age and sex of each individual in addition to test year and country. Estimated heritabilities for traits included in the selection criteria ranged from 0.22 to 0.46 for conformation traits and from 0.20 to 0.58 for riding ability traits (Albertsdóttir *et al.*, 2008).

Estimated genetic correlations between traits included in the selection criteria ranged from -0.22 to 0.88, where the highest correlations were within the riding ability traits (Albertsdóttir *et al.*, 2008).

Currently, developments in the breeding evaluation are being made regarding adding information about competition records and test status (Albertsdóttir *et al.*, 2012). Competition traits are reported to have moderate estimated heritabilities (Albertsdóttir *et al.*, 2007) and high genetic correlations with the breeding field test traits, especially with the riding ability traits (Albertsdóttir *et al.*, 2008). Test status describes whether a horse was assessed in a breeding field test or not. Approximately 12% of registered horses participate in these field tests, 5% of stallions and 19% of mares. Estimated heritabilities of test status were high (0.51-0.67) and the trait is assumed to account for preselection in the breeding field test traits. Genetic correlations between test status and breeding field tests traits reflect the weight of the traits in the selection criteria, indicating that most breeders are complying with the official breeding goal (Albertsdóttir *et al.*, 2011). Integration of the trait to the genetic evaluation reduces selection bias and increases accuracy of the estimation.

Information on estimated breeding values is available in the WorldFengur database. No regulations apply to how the breeders should perform their selection but considerable genetic gain according the official breeding goal has been achieved over the last two decades (Sigurðardóttir, 2012).

2.1.4 Breeding field tests

Standardised breeding field tests are performed yearly in various locations in Iceland, where subjective assessments for conformation and riding ability traits included in the selection criteria are made by a panel of three certified breeding judges. The combination of breeding judges in the judging panel varies between different breeding field tests. Horses need to have reached the age of four years old to participate in full assessment and can be assessed repeatedly after that. All horses must be registered and individually identified to participate in breeding field tests and parentage proven with a DNA analysis (FEIF, 2016a). The procedure starts with taking body measurements from the horse which is subsequently used as an aid for conformation assessments. Upon completion of conformation assessments, the first assessment for riding ability is performed outside where the horse is ridden alone five times back and forth on a level, straight track in front of the judging panel. At the end of each breeding field test when all horses have been assessed, a second assessment for riding ability is performed. This time the horses are ridden three times in each direction on the same track in a group of 2-3 horses, where they are given the opportunity to improve their performance and raise the score for riding ability. However, spirit and form under rider are composite traits as they are assessed based on the whole performance of all gaits and scores for these traits are therefore rarely raised during the second assessment. All traits are scored on a linear scale from 5.0 to 10.0 with half points given, where 5.0 is given for a riding ability trait not presented. Additional standardised comments are given for each assessed trait to describe certain characteristics of the trait and to substantiate the scoring. The FAI and FEIF administer the breeding field tests and ascertain that assessment procedures are standardised in order to ensure comparability between membership countries (FEIF, 2016a). All information collected at breeding field tests is available in WorldFengur.

The judging panel generally consists of three certified breeding judges who reach a joint conclusion on the assessment for each trait. The prerequisites for becoming a certified breeding judge in Iceland are a completed university degree in agricultural or animal science, a specific qualifying exam administrated by the FAI and a qualifying FEIF exam for international breeding judges. As the breeding assessments are subjective, the requirements for breeding judges to remain unbiased and exhibit refinement in their work are high. Riders performing at breeding field tests are not required to be specifically qualified for riding, but usually these riders are professional riders with great experience. Some riders specialise in breeding field test riding and ride numerous horses in these field tests and have not necessarily trained these horses before. Many of the professional riders however both train the horse before and ride it during the test. Few non-professional riders perform at breeding field tests, but in these cases they are usually riding their own horses which they have also been training. However, most breeding horse owners employ professional riders to ride their horses at breeding field tests, as the results from these tests are highly dependent on the rider being able to elicit the horse's best qualities which has a great importance for the economic value of the horse.

2.2 Spirit

The attributes of spirit are described as the willingness and disposition of the horse and how sensible and easy it is to handle (Kristjánsson, 2014). The breeding goal for spirit and the representation of the highest possible score for the trait describes a fiery, cheerful and brave horse that is extremely easy to handle and tries to please the rider at all times (FEIF, 2016a).

Spirit has been a part of the selection criteria since 1950, but until 2000 it was defined as two separate traits; temperament and willingness. The arguments for incorporating temperament and willingness into one trait were the close internal relation between the two traits, which made it difficult to assess them separately, and the fact that the interplay between them was believed to be the true value (Kristjánsson, 2001). The weight for willingness and temperament was 17.5% and 10% respectively when it was highest (Arnórsson, 2004), but the weight in the selection criteria has decreased over time and today the weight for spirit is 9% of the total score (FEIF, 2016a). Heritabilities for willingness and temperament were estimated at 0.30 and 0.23 respectively (Árnason & Sigurðsson, 2004). The estimated heritability used for the current breeding evaluation of spirit is 0.37 (Árnason & Sigurðsson, 2004). The heritability for spirit was also estimated to be 0.37 by Albertsdóttir *et al.* (2008). In a review by von Borstel (2013), the majority of heritability estimates for personality traits assessed along performance tests in various horse breeds ranged from 0 to 0.40, where the higher estimates were based on data collected at on-station tests. Spirit was shown to be moderately to highly genetically correlated to most of the riding ability traits (0.33-0.88) (Albertsdóttir *et al.*, 2008) and with a high estimated genetic correlation to the breeding field test status (0.85) introduced by Albertsdóttir *et al.* (2011), indicating a preselection of breeding horses participating in breeding field tests based on spirit. Genetic gain for spirit has been reported to be 0.42 index units (mean of the reference population is 100 index units with a standard deviation of 10) per year on average over the period of 1990-2010 (Sigurðardóttir, 2012). This was relatively high genetic gain compared to other traits reported in the same study and reflects the weight of the trait in the selection criteria. Truncation selection percentage for spirit was reported around 17% for stallions and around 80% for mares in 2010 in the same study.

The assessment of spirit is conducted during the riding ability assessment at breeding field tests. The breeding judges assess the apparent temperament and willingness of the horse through the whole performance, taking into account behaviour and capability in all gaits as well as cooperation during change of direction and speed changes. Subsequently the judges give an appropriate score according to the judging scale and describe each horse with standardised comments for prominent attributes of the trait. Obvious temperament fault displayed by a horse during the conformation assessment are taken into account when assessing spirit. For a horse to receive the highest score for spirit (9.5-10.0), it needs to display great willingness and capability in all gaits, be completely free of tension and nervousness, express excellent suppleness and cooperation towards the rider and an overall fiery and cheerful appearance. Horses that exhibit stubbornness, extreme nervousness, laziness or out of control behaviour receive the lowest scores (5.0-7.0), which is however uncommon and the score 5 is only assigned if the trait is not shown during the test, which is never the case with spirit due to its composite nature. Detailed description of the judging scale for spirit is presented in Appendix I. Of all the riding ability traits, spirit had the highest average score but the lowest standard deviation which described low variation of scores for the trait in a previous study (Helgadóttir, 2009). Scores can be raised during the second riding ability assessment if the performance of the horse improves but in order to raise the score for spirit the horse is required to perform in at least three different gaits because of the composite nature of the trait. Therefore, the frequency of raising the score for spirit has been relatively low (3%) compared to other riding ability traits (4%-20%) (Indriðadóttir, 2012).

Spirit has been one of the most controversial traits assessed at breeding field tests due to the subjective assessment method which has been suggested to favour nervous and tense horses but supple and stable horses better suited for general riders receive lower scores for spirit. This suggestion was partly supported by the findings of Brunberg *et al.* (2013) that horses with a high BLUP value for spirit seem to express stronger fear reactions, but the study was based on results from very few horses and the effect of age was not included. Discussions regarding these concerns were carried out during a symposium on horse breeding in Hvanneyri, Iceland in 2013 by general breeders and officials of the Horse Breeders Association in order to contribute to improvements of the assessment method. Ideas such as delegating the weight of spirit to the other riding ability traits and thereby removing the trait from the selection criteria were discussed. Most participants however concurred that spirit is one of the most valuable traits for a riding horse and therefore it is essential to keep the trait as a part of the selection criteria and emphasise more on rewarding supple and pleasing horses. Other ideas included specific behavioural- and temperament tests for breeding horses and registration of reason for culling in WorldFengur (The Agricultural University of Iceland, 2013). Following the symposium, the official Horse Breeders Association council initiated a project involving additional assessments made by breeding judges at breeding field tests pertaining to temperamental suppleness of breeding horses (Horse Breeders Association council, 2013).

2.2.1 Suppleness traits

The assessment method for suppleness traits within the aforementioned project was prepared by Þorvaldur Kristjánsson, the chairman of the Horse Breeders Association in Iceland, and Víkingur Gunnarsson, an experienced breeding judge, to be carried out at breeding field tests in Iceland in 2014 and continued for an indefinite period of time. The assessment method is based on assessing five traits pertaining to the suppleness of the horse; head carriage, rein contact, top line, cooperation and nerve strength. The traits are assessed on a linear scale from 1 to 7, where 7 is the highest score for each trait. In addition, a score for the overall suppleness of the horse is assessed on the same scale, based on the attributes of the five traits rather than a summarisation of the scores. Details of the judging scale for score for overall suppleness is presented in Appendix II. The assessments are carried out by breeding judges simultaneously with routine procedures during the first assessment of riding ability traits at breeding field tests and cannot be raised during the second assessment.

Assessment for head carriage involves how stable the head carriage is. A horse with a stable, well-carried head receives the highest score for this trait and a horse with unstable or stiff and heavy head carriage and/or frequent head tossing receives the lowest score. When assessing rein contact the emphasis is on a light and supple rein contact between rider and horse. A horse displaying this quality receives the highest score for rein contact but where rein contact is stiff and heavy and the horse's response to the rein aids is negligible the lowest score is given. Assessment for top line involves assessing how able and willing the horse is to carry it self correctly. A horse displaying correct collecting ability with supple and elastic top line receives the highest score but a lack of collection and stiff top line results in the lowest score. When assessing cooperativeness of the horse the emphasis is on how easily the horse responds to the rider's aids for example when turning around on the track and during speed or gait changes. A horse that is eager, ready and willing to respond with lightness receives the highest score for cooperation. A horse that is uncooperative and aggressive or even displaying out of control behaviour receives the lowest score for cooperation. Assessment for nerve strength involves assessing how composed and courageous the horse is. A composed and courageous horse displaying determination in all its tasks receives the highest score for nerve strength, but a tense, nervous horse displaying extreme fearfulness receives the lowest score.

The highest score for overall suppleness describes a horse that is extremely supple, light and cooperative in every aspect. It is apt, alert, focused, courageous and completely free of tension. Rein contact is light and supple in all gaits and the horse is able and willing to carry it self correctly with a supple top line and free of stiffness. A horse that receives the lowest score for overall suppleness is described as out of control, extremely stubborn or attempting to bolt or rear/buck, or a horse that is not able to finish the test due to commotion or decisive cooperation faults during the assessment. If a horse expresses obvious temperament faults during the conformation assessment (for example lack of cooperation, tension, coldness, nervousness), it is considered during the suppleness assessment in the riding ability assessment. The score for overall suppleness is therefore an additional independent score based on the attributes of the five traits rather than a summarisation of their scores.

Following the breeding field tests in 2014, genetic analyses for the suppleness traits were carried out (Birgisdóttir, 2015). Estimations were based on 1162 assessments including repeated assessments and the model was adjusted for fixed effects of field test and the interaction between age and sex. Estimated genetic parameters are presented in Table 2. Genetic correlations within the traits were high, ranging from 0.72 to 1.00, where head carriage, top line and cooperation were estimated to have the highest genetic correlation to the other traits. It was concluded that these traits describe the same genetic variances as rein contact, nerve strength and suppleness, and therefore the suppleness assessment was changed before upcoming breeding field tests in 2015 to only include rein contact, nerve strength and overall suppleness. However, the judging scale and assessment methods for these traits remained the same between the two years. As the suppleness assessment procedure is still being developed, the information collected pertaining to this part of the riding ability assessment is not available in WorldFengur.

Table 2. Heritabilities (h^2), additive genetic (σ^2_a) and permanent environmental (σ^2_{pe}) variances (standard errors as subscripts) estimated in univariate analysis for suppleness traits assessed by judges at breeding field tests in Iceland in 2014 (Birgisdóttir, 2015)

Trait	$\mathbf{h^2}$	$\boldsymbol{\sigma^2_a}$	$\boldsymbol{\sigma^2_{pe}}$
Head carriage	0.08	$0.11_{0.07}$	$0.00_{0.12}$
Rein contact	0.05	$0.08_{0.08}$	$0.23_{0.12}$
Top line	0.32	$0.39_{0.13}$	$0.00_{0.14}$
Cooperation	0.06	$0.08_{0.08}$	$0.09_{0.12}$
Nerve strength	0.00	$0.00_{0.04}$	$0.24_{0.09}$
Suppleness	0.10	$0.11_{0.07}$	$0.08_{0.11}$

2.3 Temperament assessments in other horse breeds

Temperament traits are complex multifactorial characters (Andersson & Georges, 2004), influenced by number of genes influencing other traits simultaneously via genetic linkage and pleiotropy. Few studies have succeeded in discovering genes contributing to differences in equine temperament, but one of the recent findings is the link of the dopamine D4 receptor gene to traits such as vigilance and curiosity in horses (Momozawa *et al.*, 2005b). For the time being, breeding strategies for temperament traits in horses must therefore rely on phenotypically expressed behaviour traits, which correspondingly need to be genetically determined in order to gain genetic progress.

2.3.1 Temperament traits assessed during performance tests

Given the close relation between a horse's temperament and its capability and performance, the majority of sport horse breeding associations worldwide include temperament assessments in their breeding programs, carried out simultaneously with assessments of performance traits such as basic gaits and jumping ability (Koenen *et al.*, 2004; von Borstel, 2013). As reported in a review on equine studies (von Borstel, 2013), these temperament assessments included various different traits such as character and willingness to work, but the trait temperament was almost universally included, indicating a great importance of the trait. It was commonly defined based on aspects of fear reactivity. The same review reported that estimated heritability for the trait temperament in various horse breeds derived from assessments given in field and/or station tests ranged from 0.03 to 0.76 where most values were within the range of 0.10 to 0.40. Estimated heritabilities for the traits character and willingness to work ranged from 0.06 to 0.27 and 0.10 to 0.29, respectively. Estimated heritabilities for other, less frequently assessed traits such as constitution (0.10-0.35) in Warmblood horses, docility (0.02) in Haflinger horses and handleability (0.09-0.22) in Haflinger and Southern German Draught horses were reported in the same review. Southern German Draught horses are also assessed for the traits nerve strength and concentration with estimated heritability of 0.14 and 0.17 respectively. Most of the aforementioned heritabilities were estimated using multivariate animal models including fixed effects such as test age or birth year, sex, test year and place of testing, or the combination between two of them. Heritability of reactivity assessed during conformation evaluation at field tests for Danish warmblood horses was estimated at 0.17 using an animal model with no fixed effects (Rothmann *et al.*, 2014a).

2.3.2 Temperament traits assessed in specific behaviour tests

Few studies have focused on heritability estimates for temperament traits included in specific behaviour observations taken during tests or everyday situations as reported in a review by von Borstel (2013). The most common traits assessed in these studies were reactivity and emotionality. Estimated heritabilities for these traits were reported moderate to high for horse breeds such as Selle Francais sport horses (0.24-0.81) in a novel object test, arena test and bridge test; Warmblood sport horses and mixed breeds of ponies in Germany (0.26-0.40) in a riding novel object test; Malopolski riding horses (0.39-0.90) in a novel object test and Franches Montagnes (0.17-0.26) in a riding or driving novel object test and bridge test. Heritability estimates of other traits such as handling traits (0.23-0.28) from a routine veterinary examination of Thoroughbred racehorses; attention to environment (0.21), attention to rider (0.18) and intensity of rider‘s aids (0.17) in a riding novel object test for German Warmblood horses and ponies; and trainability (0.10) rated by owners for Standardbred horses have also been reported (von Borstel, 2013). The study on German warmbloods and ponies reported moderate to high genetic correlation between reactivity in a novel object test and basic gaits (0.28-0.72) and rideability (0.67). Number of tested horses in these studies varied between 127 and 703, except the study on the Thoroughbred racehorses which included 4452 tested horses. These studies indicate that temperament traits assessed during specific behaviour tests could be implemented into breeding programs, but the low number of studies and low number of tested animals included in most of them suggests that further research is needed. Economic aspects and practicability need to be taken under consideration as well. Furthermore, Seaman, Davidson and Waran (2002) reported a great variety and inconsistency over time in responses by horses to three different behavioural tests in three trials; arena test, response to a person and response to an object. Open-field arena test was the only test where the horses were found to behave consistently over the three trials and was deemed the only type of test indicating some core factor of temperament. It could not however be used to make a prediction of behaviour in the two other tests and vice versa.

2.3.3 Limitations of temperament assessments implemented in breeding programs

As mentioned before, the majority of sport horse breeding associations include assessments of temperament traits in their breeding programs. However, only few breeding associations use the obtained information from temperament assessments to estimate breeding values to use in selection, because there is reluctance to base selection on subjective assessments obtained during performance tests (von Borstel, 2013). Subjectivity of assessments methods, lack of appropriate guidelines for assessment of personality traits, lack of reliance on existing guidelines, inadequate definitions of personality traits within the guidelines, and economic competitions are common problems within breeding programs (von Borstel *et al.*, 2011b; von Borstel, 2013). These problems can result in an inflation of scores, bias and an undesirably low variability in scores as reported by Pasing and von Borstel (2012), but increase in scores for personality traits for German stallions in performance tests were detected beyond scores for other performance traits which could not be explained with genetic gain or improvements in training techniques. Horse breeding authorities have therefore made attempts to improve assessment procedures by developing more objective assessment methods, for example by implementing physiological measures such as heart rate and heart rate variability measures (Visser *et al.*, 2002; Visser *et al.*, 2003; Górecka-Bruzda *et al.*, 2011a; von Borstel *et al.*, 2011a, 2011b, 2012).

2.3.4 Towards a more objective assessments of temperament traits

Visser *et al.* (2002) reported that heart rate and heart rate variability used along novel object and handling test quantify certain aspects of a horse‘s temperament and von Borstel *et al.* (2011b) reported several associations between behavioural and physiological observations, and suggested that they should be considered when redesigning current guidelines for assessment of personality traits during breeding horse performance tests. Von Borstel *et al.* (2011a) concluded that a rider or handler influenced but did not completely mask the behaviour of the horse in personality tests and therefore personality tests that resemble the practical circumstances most closely should be chosen for valid assessments. Furthermore, von Borstel *et al.* (2012) reported that temperament tests using novel stimuli presented to a horse under a rider may be practical and valid tool for improving the current assessment of temperament traits during performance tests.

2.3.5 Score ratings by riders and judges

Few studies have focused on the aspect of comparing score ratings made by riders or judges and behaviour observations. Visser *et al.* (2003) studied the response of horses in behavioural tests and estimated correlations with temperament assessments made by the riders. It was concluded that large panel of assessors could agree upon a horse’s temperament, and objective measures from behavioural tests correlated significantly with temperament assessments made by a panel of assessors, but not with novel object test. Von Borstel *et al.* (2011b) conducted an experiment where specific behavioural traits where observed during a training phase and performance test and compared the observations to assigned scores for temperament traits assessed by judges at the same time. Behavioural traits such as head-tossing, stumbling, the horses’ head posture, horse-induced change in gait and the rider’s use of voice had significant influences on scores for several different temperament traits. This overlap suggested that the judges did not clearly distinguish between the different temperament traits but rather used their overall impression of the horse to assign the scores. Rothmann *et al.* (2014b) investigated the association between scores for reactivity (assessed with behaviour observations during conformation assessment at field tests) and scores given for rideability and performance by judges at field tests. Results indicated that less reactive horses received higher scores in free jumping and for rideability and it was concluded that reactivity measures are possible in practical situations with further development and adaptations of behaviour scores.

2.3.6 Questionnaire surveys in animal studies

Generally, questionnaires are considered subjective but a number of studies have focused on the aspects of assessing temperament traits in horses by conducting questionnaire surveys for owners and/or trainers. Momozawa *et al.* (2003) compared heart rate and behaviour in a behaviour test and scores from a questionnaire answered by caretakers and concluded that a questionnaire survey could be an effective way to assess temperament traits in horses. However, in aforementioned study by Seaman *et al.* (2002) where variety and inconsistency over time in responses by horses to artificial behaviour tests was reported, a questionnaire completed by caretakers was used to validate the test results and no relationship was found between the responses in the tests and the ratings given by the caretakers. Few studies have aimed at developing a validated questionnaire for the purpose of assessing behaviour and temperament traits in horses and dogs (Momozawa *et al.*, 2005a; Hsu & Serpell, 2003; Arvelius *et al.*, 2014). Arvelius *et al.* (2014) studied the heritability of everyday life temperament in dogs with records from questionnaires answered by owners, and estimated the heritability ranging from 0.06 to 0.36 and high genetic correlations with traits in temperament tests these dogs were subjected to, indicating reliability of the assessment method to a certain extent. In a previously discussed study by Rothmann *et al.* (2014b) association between reactivity assessed with behaviour observations and temperament traits (nervousness, agreeableness and trainability) scored by owners and/or trainers in a questionnaire survey were investigated. Results indicated that highly reactive horses were assessed more nervous by their owner and/or trainer but no correlation was detected between reactivity and agreeableness or trainability. Axel-Nilsson *et al.* (2015) conducted a questionnaire survey with the aim of defining a list of behavioural traits in horses which could be of use in optimising a good match between horse and rider, and received answers from over 2800 participants, indicating a great interest in the topic and good presentation of the survey.

2.3.7 *Culling due to temperament faults*

The procedure of culling or retiring horses due to temperamental disorders has only been reported in five horse breeds; the Icelandic horse (Björnsdóttir *et al.*, 2003; Sigfússon, 2003), Swedish Warmblood and Coldblood draught horses (Wallin *et al.*, 2000) and Thoroughbred and Standardbred race horses (Hayek et al., 2005). As said before, the frequency of culling because of temperamental disorders in Icelandic horses has been reported to be 16% of all culled horses, based on a sample of 201 horses (Sigfússon, 2003) and 98 horses (Björnsdóttir *et al.*, 2003). However, when culling reasons were distinguished between voluntary and involuntary culling, temperamental disorders were reported to account for 31% of voluntary culling and the mean age of horses culled due to temperamental disorders was 7.1 years (Sigfússon, 2003). Sigfússon (2003) also reported that about 1% of culled horses were culled due to old age. The study by Björnsdóttir *et al.* (2003) only included horses in the age range 6-12 years and the mean age was not reported. The reported frequency of culling due to temperament faults in Swedish Warmblood and Coldblood draught horses was 1% and 23%, respectively. The Coldblood horses had a median length of life approximately 18 years, which was 3 years longer median length of life then the Warmblood horses. The difference was attributed to the area of use, as most of the Coldblood horses were draught horses. The majority of the Warmblood horses were used for sport and were most frequently culled due to diseases of the musculoskeletal system. Temperamental disorders were reported to be the cause of retirement for 6.4% of Thoroughbred and Standardbred race horses in Australia, which were at high risk of entering slaughterhouses.

3 Aim of the thesis

The specific aims of this thesis were:

- To estimated genetic parameters for temperament traits assessed by breeding judges for all horses presented at all breeding field tests in Iceland in 2014 and 2015.
- To compare assessments made by riders and judges for temperament traits of the same horses presented at selected breeding field tests by estimating genetic parameters for the assessments made by both groups.
- To investigate the difference between assessments made by riders and judges by identifying possible underlying relationship between traits assessed by both groups.
- To estimate genetic parameters for general temperament traits expressed by horses in their everyday environment assessed by owners, breeders and/or trainers.
- To compare general temperament traits assessed by owners, breeders and/or trainers to temperament traits assessed by judges at breeding field tests.
- To estimate frequency of culling due to temperament faults in the Icelandic horse population.

4 Materials and methods

The thesis includes three different investigations. In the first part genetic parameters were estimated for temperament traits assessed by breeding judges and breeding field test riders (BFT riders) at selected breeding field tests in Iceland during the summer of 2014 and 2015, along with genetic parameters for temperament traits for all horses assessed by breeding judges at breeding field tests in Iceland in 2014 and 2015. In the second part genetic parameters were estimated for general temperament traits observed at home, based on assessments made by owners, breeders or/and trainers. The horses included were tested at breeding field tests in Iceland during the summer of 2014 and 2015. In the third part of the thesis the frequency of culling due to temperament faults was estimated, based on answers from the owners of horses culled during the period of September 2014 to January 2015.

4.1 Materials

4.1.1 Part I

All assessed horses

All available assessments made by breeding judges in breeding field tests in Iceland in 2014 and 2015 for spirit and suppleness traits (head carriage, rein contact, top line, cooperation, nerve strength and overall suppleness) were provided by the FAI. The score for spirit was the final score after the second ridden assessment. Assessments for suppleness traits were unavailable from two breeding field tests.

In total 2537 assessments for total score were made in 2014 and 2015 in Iceland for 1932 different horses consisting of 72.2% mares and 27.8% stallions or geldings. The assessments were performed at 30 different breeding shows in total and the horses were ridden by 240 different riders. All the horses presented at these breeding field tests had known parentage apart from five individuals which had an unknown mother. The mean estimated breeding value (EBV) for total score of conformation and riding ability for all individuals assessed at breeding field tests in Iceland 2014 and 2015 had a mean EBV of 111 with the range of 74 to 130. Sires and dams of those individuals had a mean EBV of 117 and 105 respectively. Breeding values were estimated by the FAI in the autumn of 2015.

Sample group

Data was randomly collected from BTF riders at four different breeding field tests at Gaddstaðaflatir, Hella in Iceland in 2014 and 2015, and concluded in a sample of 451 assessments for 440 different horses, which represents 67.0% of the horses assessed for total score at the field tests at Gaddstaðaflatir and 22.8% of all horses assessed for total score in various breeding field tests in Iceland 2014 and 2015. Data collected from the riders included assessments for spirit and the suppleness traits based on the performance of the horse during the first ridden assessment. Assessments made by breeding judges and riders were therefore conducted for the same performance of each horse. In 2015, data on suppleness traits only included assessments for rein contact, nerve strength and overall suppleness as explained in chapter 2.2.1. The riders used the same judging scales as breeding judges for both spirit (linear scale 5-10 with half points given) and the suppleness traits (linear scale 1-7). In addition, the riders were asked to give information if they had been training the horse themselves before the test and if so, the amount of training the horse had been subjected to (measured in months of training). The riders were also asked to inform if they were the owner of the horse as well. The questionnaires that the riders were subjected to each year are included in Appendix III. A total of 80 different riders participated in the study. Riders were divided into 3 groups based on experience and total number of horses ridden in breeding field tests in 2014 and 2015, in order to investigate effects of the riders in the genetic analyses. Group 1 included 26 riders that showed less than 6 horses each, group 2 included 31 riders with 6-20 horses and group 3 included 23 riders with more than 20 horses. One rider was moved from group 1 to group 2, based on his extensive experience as a riding instructor. Assessments for spirit made by riders were missing for eight horses and 16 horses for overall suppleness. Assessments given in .5 by few of the riders for traits scored on the 7 point scale were rounded up to the next integer. Few riders gave the score 5 for spirit which was treated as missing data, as the score 5 is only used for traits not displayed during the field tests.

Assessments made by breeding judges for the same horses were provided by the FAI, except for the trait spirit which was obtained from WorldFengur after the first ridden assessment. Assessments for suppleness traits made by judges in 2014 were missing for five individuals.

The sample group consisted of 78.4% mares and 21.6% stallions or geldings where all horses had known parentage except two individuals which had an unknown mother. The mean EBV for total score of conformation and riding ability for the horses included in the sample was 113 with the range of 93 to 126. Sires and dams of those individuals had a mean EBV of 117 and 106 respectively. Breeding values were estimated by the FAI in the autumn of 2015.

Details of analysed data are presented in Table 3, for all horses assessed at breeding field tests in Iceland in 2014 and 2015 and the sample where riders and judges assessed the same horse.

Table 3. Details of analysed data for horses assessed in breeding field tests in Iceland in 2014 and 2015, both for all horses assessed in both years and the sample group

	All assessed		**Sample**	
	2014	**2015**	**2014**	**2015**
Total number of assessments	1523	1014	231	220
Number of individuals assessed	1101	831	225	215
Number of mares:				
4 years	71	34	24	9
5 years	178	158	42	48
6 years	243	181	51	54
7 years and older	297	233	48	69
Total	789	606	165	180
Number of stallions/geldings:				
4 years	60	45	12	4
5 years	107	72	23	11
6 years	89	71	15	16
7 years and older	56	37	10	4
Total	312	225	60	35
Number of different riders	186	174	50	62
Number of different breeding field tests	14	16	2	2
Number of known sires	282	246	112	120
Average no. of assessed offspring per sire*	3.9	3.4	2.0	1.8
Number of known dams	946	750	213	203
Average no. of assessed offspring per dam*	1.2	1.1	1.1	1.1

*Average number of offspring assessed in 2014 and 2015 in the sample group where riders and judges assessed the same horses and in the total group of horses assessed in these years.

4.1.2 Part II

Data was collected through an online questionnaire survey aimed at owners, breeders and/or trainers of all horses assessed for riding ability at breeding field tests in Iceland during the summer of 2014 and 2015. The survey was available on the webpage for WorldFengur database (www.worldfengur.com) during the period of October 12th to November 29th 2015. The survey was made accessible online with assistance from the IT department at the FAI and advertised through horse specific online media in Iceland as well as through emails sent to the owners who had registered email addresses in WorldFengur.

The questionnaire survey included questions about training period and training level of the horse when subjected to a breeding field test and general temperament traits such as reactivity to novel objects and sound, reactivity to disturbance in training, behaviour towards people and other horses, behaviour while being broken in and in training and general cooperation in training. Respondents were also asked to compare the score received for spirit at breeding field test to the general temperament of the horse as they perceived it, rank temperament traits according to importance and give information about the intended future role of the horse. The questionnaire was constructed with help from Icelandic leisure riders and a group of experts experienced with Icelandic horses, including breeding judges, professional trainers and riding instructors. Table 4 presents a short description of each question and the possible answers or scores. Questions 1-16 were scored on a linear scale, where 4 was the highest score, except for question 2 where one of three levels of training were chosen. If respondents answered "Unknown" it was treated as missing data. In question 17 respondents were able to rank various temperament traits where 1 was the most important trait and 7 the least important trait. In question 18, which pertained to the future role of the horse, the respondents were able to give several answers per horse. All questions were followed by a short description in order to aim for more standardized answers and decrease the risk of bias caused by subjective interpretation. If the horses had been subjected to a breeding field test in both 2014 and 2015, the respondents were able to give answers for both years based on the performance of the horse and characteristics at that time. The complete questionnaire is presented in Appendix IV.

Table 4. A short description of the questions included in the questionnaire survey and the possible answers or scores

Description	**1**	↔	**4***
Q1 Time in training	<12 months		>36 months
Q2 Training level*	Level 1		Level 3
Q3 Reaction to a novel object	Very nervous		Very calm
Q4 Action when seeing a novel object	Gets scared and flees instantly		Approaches instantly
Q5 Reaction to novel/loud sound	Very nervous		Very calm
Q6 Reaction to temporary isolation	Very nervous		Very calm
Q7 Reaction to new environment	Very nervous		Very calm
Q8 Behaviour towards humans	Very nervous		Very calm
Q9 Behaviour towards other horses	Very aggressive		Very friendly
Q10 Reaction to disturbance while being trained	Very unassured		Very assured
Q11a Reaction to new aids while being broken in	Very unapt		Very apt
Q11b Reaction to new aids while being broken in	Very tense		Very relaxed
Q12a Behaviour while being trained	Very uncooperative		Very cooperative
Q12b Behaviour while being trained	Very tense		Very relaxed
Q13 Behaviour while being ridden away from stable	Very unwilling		Very willing
Q14 General cooperation in handling	Very uncooperative		Very cooperative
Q15 Behaviour predictability	Very unpredictable		Very predictable
Q16a Consistency with the score for spirit	Very bad		Very good
Q16b Inconsistency explained	Overestimated or underestimated		
Q17 Ranking traits according to importance	Ranking 1-7		
Q18 Intended future role	Multiple options		

* Question 2 about training level only included three options (level 1, 2 or 3) and was therefore scored on a 3 point scale.

Data collected from the questionnaire survey concluded in 343 assessments but as four horses had assessments for both years only the assessments associated with the more recent breeding field test assessment were kept for these horses. The data used in the analysis therefore represented assessments for 339 horses, or 17.5% of all horses assessed for riding abilities in breeding field tests in Iceland 2014 and 2015. The dataset comprised 71.1% mares and 28.9% stallions or geldings. All assessed horses had known parentage. The mean EBV for total score of conformation and riding ability for the horses assessed in the survey was 113, ranging from 89 to 130, and the mean EBV for spirit was 112, ranging from 91 – 130. Mean accuracy of EBVs ranged between 70%-94%. Breeding values were estimated by the FAI in the autumn of 2015. Details of analysed data are presented in Table 5.

Table 5. Details of analysed data obtained from the survey about general temperament traits

	2014	2015
Number of individuals assessed	151	188
Number of mares:		
4 years	7	13
5 years	28	42
6 years	30	33
7 years or older	39	49
Total	104	137
Number of stallions/geldings:		
4 years	6	8
5 years	13	17
6 years	15	19
7 years or older	13	7
Total	47	51
Number of different respondents in survey	96	111
Number of sires	86	90
Average no. of assessed offspring per sire*	1.8	2.1
Number of dams	142	176
Average no. of assessed offspring per dam*	1.1	1.1

*Average number of offspring assessed in 2014 and 2015 for the sample obtained from the survey.

There were 178 different respondents that participated in the survey. Figure 1 represents the role of the respondent in relation to the horse in the survey. Respondents were allowed to mark multiple options when describing the role. Most of the respondents were the trainers, owners and/or breeders of the horse they assessed. 46% of the respondents were both breeders and owners of assessed horses, 35% were both owners and trainers, 30% were both breeders and trainers and 29% both trainers and BFT riders. 26% of the respondents were breeders, owners and trainers of the assessed horses.

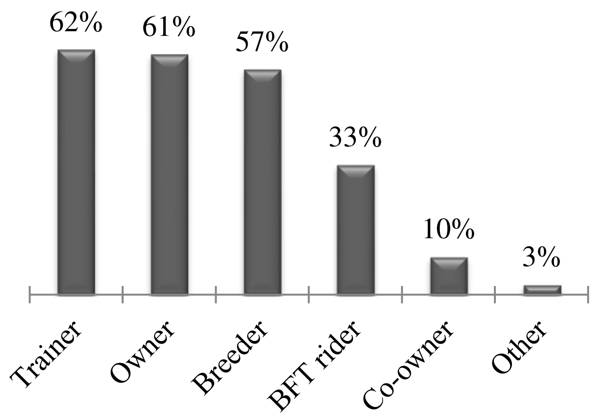

Figure 1. The role of the respondents in relation to the horses assessed in the survey.

Figure 2 presents the distribution of horses according to time in training and training level they had reached at the time when assessed at a breeding field test, estimated within each age group. The younger horses had received the least time in training measured in months but the time in training varied for the older horses. The training level gradually increased with age. Mean EBV for total score of conformation and riding ability was highest for horses that had reached level 3 of training (115) and lowest for horses in level 1 (111). Mean EBV for spirit was also highest for horses in level 3 of training (114) and lowest for horses in level 1 (110).

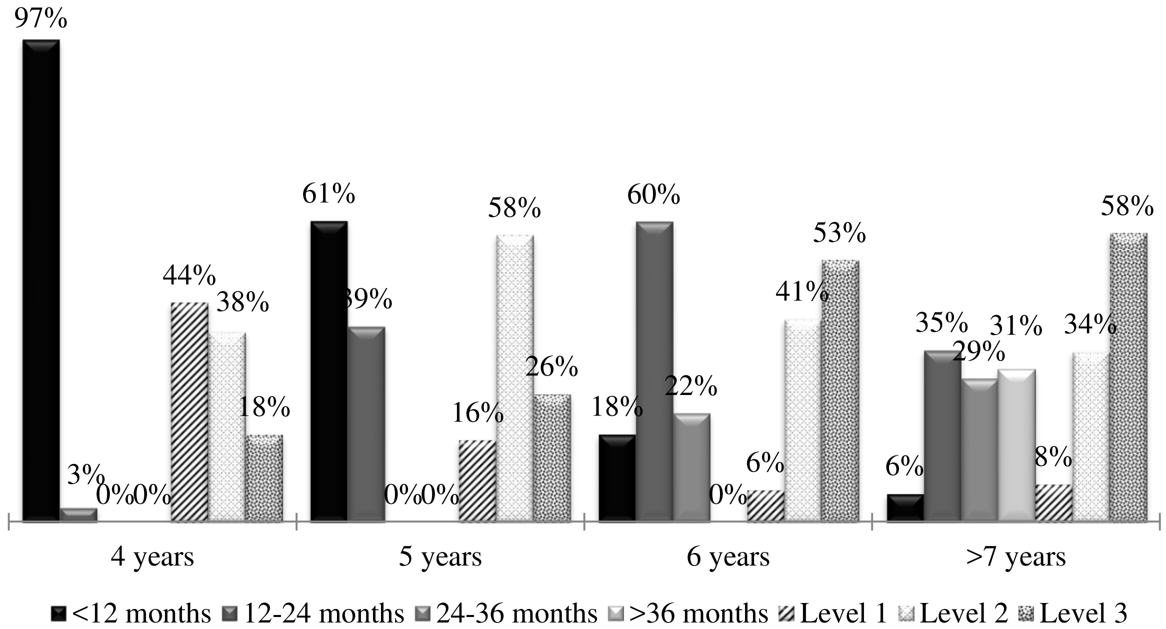

Figure 2. Distribution of horses according to time in training and training level within each age group for the horses assessed in the survey.

Assessments for temperament traits made by judges at breeding field tests in 2014 and 2015 for most of the horses assessed in the survey were available from Part I and were used for comparison. If horses had more than one assessment from a breeding field test within each year, the most recent assessment was used.

4.1.3 Part III

A list of all culled horses during the time period of September 2014 to January 2015 was provided by the FAI and included a total number of 2277 horses. Culled foals born in 2014 were not included in the list. Data collection was performed manually by contacting the owners of these horses by phone and asking them to give information about the reason for culling. In addition, the owners were asked to give information about the main role of the horse before it was culled. Due to the high number of culled horses during the time period, a decision was made to focus on horses culled in November, which had the highest number of culled horses. It is custom in Iceland to start breaking in young horses in September, and by November many owners have selected which horses will be subjected to further training based on performance and economic reasons. The data on horses culled in November should therefore partially reflect these "cut downs". Few of the owners that were contacted also had horses culled during the other months of the time period, and were asked to give answers for them simultaneously.

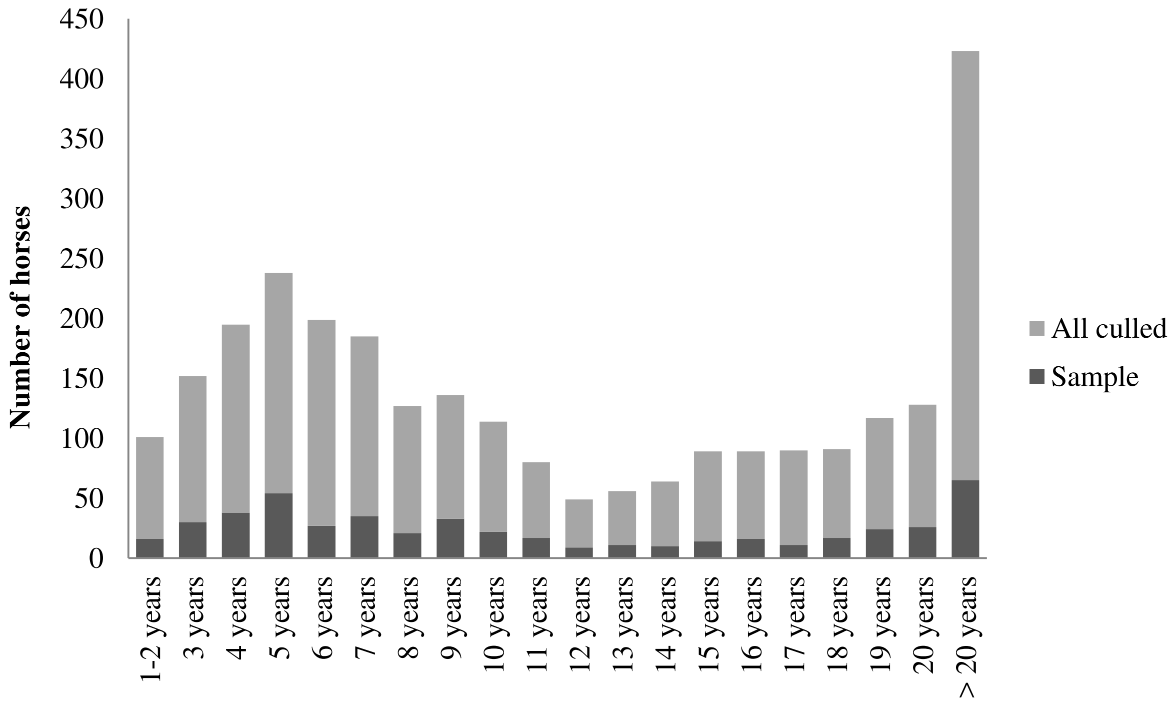

Figure 3. Age distribution for horses culled in Iceland during the period of September 2014 to January 2015, both for the sample and all horses culled during the period.

The data collection concluded in answers from 173 owners for 512 horses, where mares represented 56.1% and stallions or geldings 43.9% of the dataset. The dataset represents 22.5% of all horses culled during the time period of September 2014 to January

2015. Age distribution of culled horses is presented in Figure 3, for both the sample and all horses culled during the time period, and displays a similar distribution of both groups.

4.1.4 Pedigree data

Pedigree data was provided by the FAI and consisted of a file containing information for 241,835 horses born in Iceland, including 132,314 mares and 106,330 geldings and stallions; the sex of 3,191 individuals was unknown. The oldest individual in the pedigree was born in 1860 and the youngest in 2015. The data included individual identity number and identity numbers of both father and mother. Each identity number included the birth year and sex of the individual and a unique serial registration number. Both parents were known for 186,450 individuals.

4.2 Methods

Statistical analyses were carried out using the SAS package (SAS, 2016) and Microsoft Excel (2010) was used to produce tables and figures. Estimations for genetic parameters were carried out using the DMU software package for analyses of multivariate mixed models (Madsen & Jensen, 2013).

4.2.1 Part I

Mean, standard deviation, range, skewness and kurtosis were calculated using UNIVARIATE procedure in SAS to describe the variation of assessments made for each trait by judges and riders. Anderson-Darling test was used to ascertain normal distribution of the assessments ($p<0.005$), but as all the assessments were normally distributed no further measures needed to be taken. A general linear model (GLM procedure in SAS) was used to analyse the variance of assessments, testing sex, age, rider group, field test and test year for significance as fixed effects. The effects of field test and the interaction between age and sex were found to be the effects significantly influencing the largest number of traits ($p<0.05$). The coefficient of determination (R^2) showed that 4% to 9% of the variance could be ascribed to the model including these tested effects. Thus the final models chosen for estimation of genetic parameters for assessments made for all horses in Iceland 2014-2015 (model 1) and traits in the sample (model 2) were:

$$Y_{ijk} = \mu + \text{age\_sex}_i + \text{field test}_j + \text{individual}_k + \text{pe}_k + e_{ijk} \quad (1)$$

$$Y_{ijk} = \mu + \text{age\_sex}_i + \text{field test}_j + \text{individual}_k + e_{ijk} \quad (2)$$

Where:

Y = assessment for the k^{th} horse
μ = sample mean value for the traits
age_sex$_i$ = fixed effect of interaction between age and sex, stallions/geldings or mares (1, 2) in four age groups (4y, 5y, 6y and 7y and older)
field test$_j$ = fixed effect of field test (four different tests for the sample and 30 different for all assessments)
individual$_k$ = additive random genetic effect of the k^{th} horse ~ ND (0, σ^2_a)
pe$_k$ = permanent environmental effect of the k^{th} horse ~ ND (0, σ^2_{pe})
e$_{ijk}$ = random residual effect ~ ND (0, σ^2_e)

The fixed factors used in models 1 and 2 are the same as are currently used in the model for genetic analyses and estimation of breeding values for Icelandic horses. Details of analysis of variance in different models and significance of various fixed effects are presented in Table 1 in Appendix V.

Variance components and heritabilities were estimated with univariate models for individual traits and covariance and correlations between traits were estimated with bivariate models using average information (AI) algorithm for restricted maximum likelihood. Convergence criterion was set for the norm vector $<10^{-7}$, but for some bivariate analyses the criterion $<10^{-4}$ had to be used to get convergence. Heritabilities (h^2) were calculated as $h^2 = \sigma^2_a/\sigma^2_p$ where σ^2_a is additive genetic variance and σ^2_p is phenotypic variance (the sum of additive genetic (σ^2_a), permanent environmental (σ^2_{pe}) and residual (σ^2_e) variances).

The traits assessed by judges and riders for the same horses were subjected to a principal component analysis (PCA), to estimate differences in assessments between judges and riders. PCA was performed by using FACTOR procedure in SAS.

4.2.2 Part II

Mean, standard deviation, range, skewness and kurtosis were calculated using UNIVARIATE procedure in SAS to describe the variation of assessments for traits included in the questionnaire survey. Anderson-Darling test was used to ascertain normal distribution of the assessments ($p<0.005$), but as all the assessments were normally distributed no further measures needed to be taken. Spearman rank-order correlations were calculated between traits included in the questionnaire and suppleness traits assessed by judges at breeding field tests (head carriage, rein contact, top line, cooperation, nerve strength and overall suppleness) by using CORR procedure in SAS.

To establish levels of inner structure for the questionnaire and to condense the questions into smaller groups, data from the questionnaire was subjected to a PCA, using FACTOR procedure in SAS. The PCA summarised separate groups of common traits which was further used for genetic analysis.

A general linear model (GLM procedure in SAS) was used to analyse the variance of assessments. The effect of the interaction between age and sex was significant ($p<0.05$) for several traits and R^2 showed that 1% to 7% could be ascribed to the model including this fixed effect. Time in training and training level were both investigated as fixed effects as well as biological traits describing the ability of the horse to develop as a response to training. When investigated as biological traits, the effect of age_sex was highly significant ($p<0.0001$) for both traits and R^2 showed that 50% of the variance for time in training and 18% of the variance for training level could be ascribed to the model including this fixed effect. When investigated as fixed effects, time in training along with age_sex had significant effect for training level ($p<0.0001$) and explained 28% of the variance of assessments, but did not have significant effect on other traits. Training level along with age_sex had significant effect for time in training ($p<0.0001$) and explained 55% of the variance of assessments. The effect of training level was significant ($p<0.05$) for several other traits and R^2 showed that 4% to 10% of the variance could be ascribed to the model including fixed effect of training level and age_sex. Thus, three different models were used for estimation of genetic parameters; model 3 for all traits included in the questionnaire survey, model 4 for all traits except for training level and model 5 only for training level:

$$Y_{ij} = \mu + \text{age\_sex}_i + \text{individual}_j + e_{ij} \qquad (3)$$

$$Y_{ijk} = \mu + \text{age\_sex}_i + \text{training level}_j + \text{individual}_k + e_{ijk} \qquad (4)$$

$$Y_{ijk} = \mu + \text{age\_sex}_i + \text{training time}_j + \text{individual}_k + e_{ijk} \qquad (5)$$

Where:

Y = assessment for the j^{th} / k^{th} horse
μ = sample mean value for the trait assessments
age_sex$_i$ = fixed effect of interaction between age and sex, stallions/geldings or mares (1, 2) in four age groups (4y, 5y, 6y and 7y and older)
training level$_j$ = fixed effect of training level (level 1, 2 or 3)
training time$_j$ = fixed effect of training time (<12, 12-24, 24-36 and >36 months)
individual$_{j(k)}$ = additive random genetic effect of the j^{th} / k^{th} horse ~ ND (0, σ^2_a)
e$_{ij(k)}$ = random residual effect ~ ND (0, σ^2_e)

Genetic variance components were estimated with univariate models for individual traits and covariance and correlations between common traits summarised by PCA were estimated with bivariate models using average information (AI) algorithm for restricted maximum likelihood. Convergence criterion was set for the norm vector $<10^{-7}$, but for some bivariate analyses the criterion had to be lowered to norm vector $<10^{-5}$.

Model 3 was used to estimate genetic parameters for all the general temperament traits and when analysed with spirit. Model 4 was used to estimate genetic parameters for all the general temperament traits (except training level) when analysed with spirit. Model 5 was used to estimate genetic parameters for training level when analysed with spirit. Details of analysis of variance of different models and significance of fixed effects are presented in Table 2 in Appendix V.

4.2.3 Part III

Frequency of culling due to temperament fault, along with other reasons for culling, was estimated with the FREQ procedure in SAS. Distribution of previous role of the horse as well as age distributions was also estimated with FREQ procedure.

5 Results

5.1 Part I

5.1.1 Descriptive statistics

All assessed horses

Mean scores assessed by judges for all horses tested at breeding field tests in Iceland in 2014 and 2015 ranged between 4.25 and 4.96 for the suppleness traits, with the lowest score for top line and highest score for nerve strength. Mean score for spirit was 8.38 with scores ranging from 6.5 to 10.0, but as said before scores below 6.5 are rarely used and the score 5 is never used for spirit due to its composite nature. Distribution of scores was close to symmetry for most traits, with a minor skewness to the left for cooperation and nerve strength. Kurtosis values indicate a remote light tailed distribution for all traits except spirit, cooperation and nerve strength, which are close to standard normal distribution. All traits tested significant for normal distribution according to Anderson-Darling test ($p<0.005$). Statistics for scores given for temperament traits assessed by judges for all horses that attended breeding field tests in 2014-2015 are presented in Table 6.

Table 6. Number of assessments, mean score ($\bar{x}$), standard deviation (S.D.), range, skewness and kurtosis for temperament traits assessed by judges for all horses assessed at breeding field tests in Iceland in 2014 and 2015

Trait	No. of assessments	$\bar{x}$	S.D.	Range*	Skewness	Kurtosis
Spirit	2537	8.38	0.44	6.5 - 10.0	0.15	-0.03
Head carriage	1171	4.52	1.20	1.0 - 7.0	-0.18	-0.33
Rein contact	2088	4.44	1.21	1.0 - 7.0	-0.10	-0.36
Top line	1171	4.25	1.15	1.0 - 7.0	0.15	-0.44
Cooperation	1171	4.92	1.18	1.0 - 7.0	-0.42	0.02
Nerve strength	2088	4.96	1.14	1.0 - 7.0	-0.37	-0.06
Suppleness	2087	4.76	1.10	1.0 - 7.0	-0.14	-0.20

*Maximum score on scale for spirit is 10.0 and minimum score is 5.0. Maximum score on scale for head carriage, rein contact, top line, cooperation, nerve strength and suppleness total is 7.0 and minimum score is 1.0.

Sample group

Mean scores for suppleness traits assessed by riders at selected breeding field tests ranged between 5.00 and 5.25 and between 4.18 and 5.08 when assessed by judges. The lowest score was for top line when assessed by both riders and judges, but the highest was for overall suppleness when assessed by riders and nerve strength when assessed by judges. The mean score for spirit assessed by riders was 8.50, with scores ranging from 6.5 to 10.0, but the mean score was 8.35 when assessed by judges with scores ranging from 7.0 to 9.5. Distribution of scores assigned by riders was remotely skewed to the left but fairly symmetric for all traits, with scores for spirit being closest to symmetry. Distribution of scores assigned by judges was less skewed and closer to symmetry for most traits. Kurtosis values indicate a distribution with moderately strong peaks for scores given by riders for spirit and overall suppleness and the weakest peak for scores given for top line, but most traits scored by judges had a remote light tailed distribution and no prominent peaks. All traits tested significant for normal distribution according to Anderson-Darling test ($p<0.005$). Statistics for scores given for temperament traits assessed for the same horses by riders and judges at selected breeding field tests are presented in Table 7.

Table 7. Number of assessments, mean score ($\bar{\mathbf{x}}$), standard deviation (S.D.), range, skewness and kurtosis for temperament traits assessed by riders and judges for horses in the sample group

Trait	No. of assessments	Mean	S.D.	Range*	Skewness	Kurtosis
Assessed by riders						
Spirit	443	8.50	0.61	6.5 - 10.0	-0.05	0.48
Head carriage	231	5.11	1.26	1.0 - 7.0	-0.55	-0.01
Rein contact	451	5.01	1.28	1.0 - 7.0	-0.48	-0.06
Top line	231	5.00	1.43	1.0 - 7.0	-0.54	-0.28
Cooperation	231	5.18	1.45	1.0 - 7.0	-0.70	0.06
Nerve strength	451	5.23	1.36	1.0 - 7.0	-0.64	0.13
Suppleness	435	5.25	1.19	1.0 - 7.0	-0.63	0.36
Assessed by judges						
Spirit	451	8.35	0.43	7.0 - 9.5	0.12	-0.17
Head carriage	226	4.38	1.16	2.0 - 7.0	0.00	-0.47
Rein contact	446	4.43	1.21	1.0 - 7.0	-0.12	-0.42
Top line	226	4.18	1.03	2.0 - 7.0	0.00	-0.22
Cooperation	226	5.00	1.15	2.0 - 7.0	-0.31	-0.31
Nerve strength	446	5.08	1.16	2.0 - 7.0	-0.37	0.00
Suppleness	446	4.82	1.11	2.0 - 7.0	-0.24	-0.20

*Maximum score on scale for spirit is 10.0 and minimum score is 5.0. Maximum score on scale for head carriage, rein contact, top line, cooperation, nerve strength and suppleness total is 7.0 and minimum score is 1.0.

Figure 4 shows the distribution of scores given for the trait spirit by riders and judges when assessing the same horses in the sample and scores given by judges for all horses assessed in 2014 and 2015. On the scale 5.0-10.0, the most common score given by both riders and judges was 8.5, but riders gave scores using a wider range. No scores were given below 6.5 by riders and below 7.0 by judges in the sample. Scores given by judges in the sample is fairly representative of scores given for all horses.

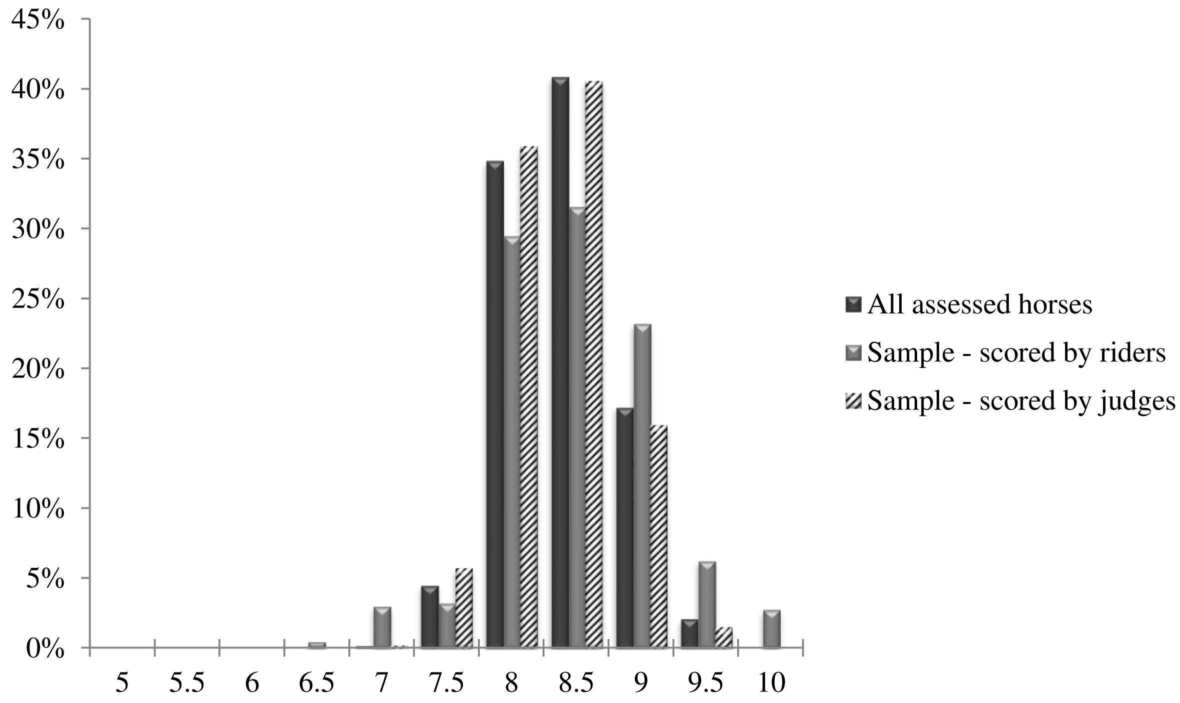

Figure 4. Distribution of scores for spirit for all assessed horses in Iceland in 2014 and 2015 and for horses in the sample group assessed by riders and judges.

Distribution of assessments made by riders and judges for each of the suppleness traits for the same horses is presented in Figure 5 and Figure 6 respectively. On the scale 1-7, 6 was the most common score given by riders for all traits except top line and cooperation, where 5 was the most common score. The most common score given by the judges was 5 for all traits except for top line, where the most common score was 4, and cooperation where the score 5 and 6 were equally common. The judges never gave the score 1 for any of the traits except for rein contact, but two individuals received that score.

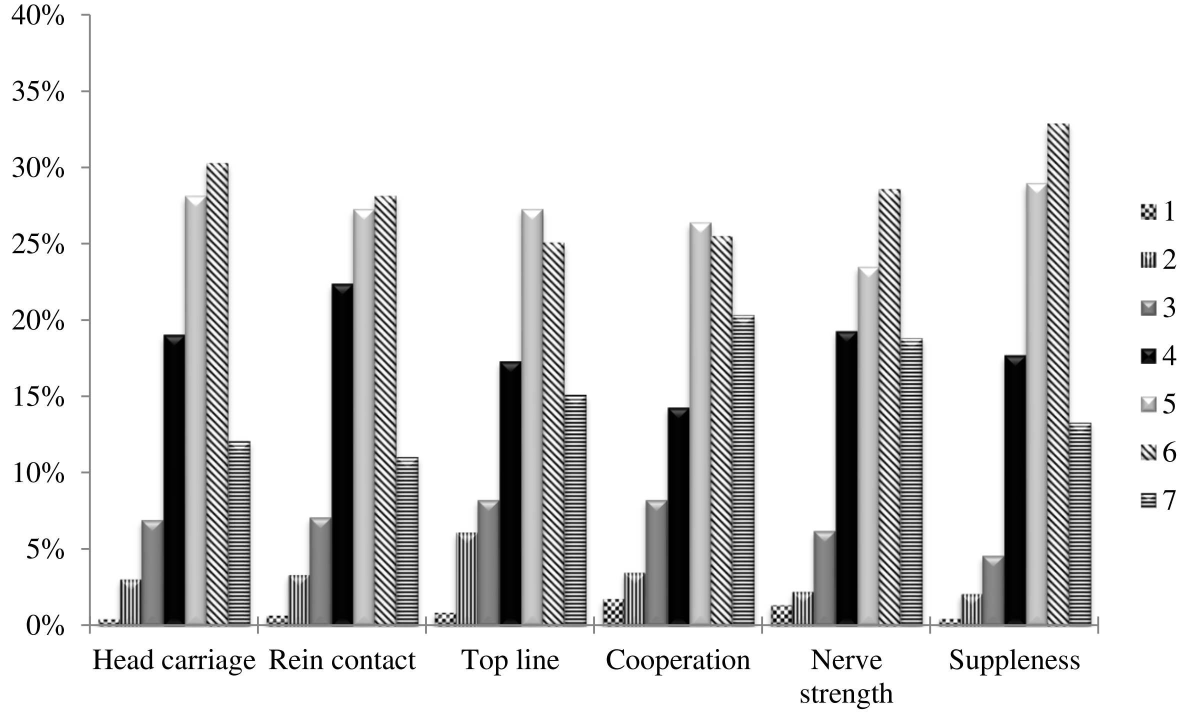

Figure 5. Distribution of scores for each of the six suppleness traits assessed by *riders* for horses in the sample group.

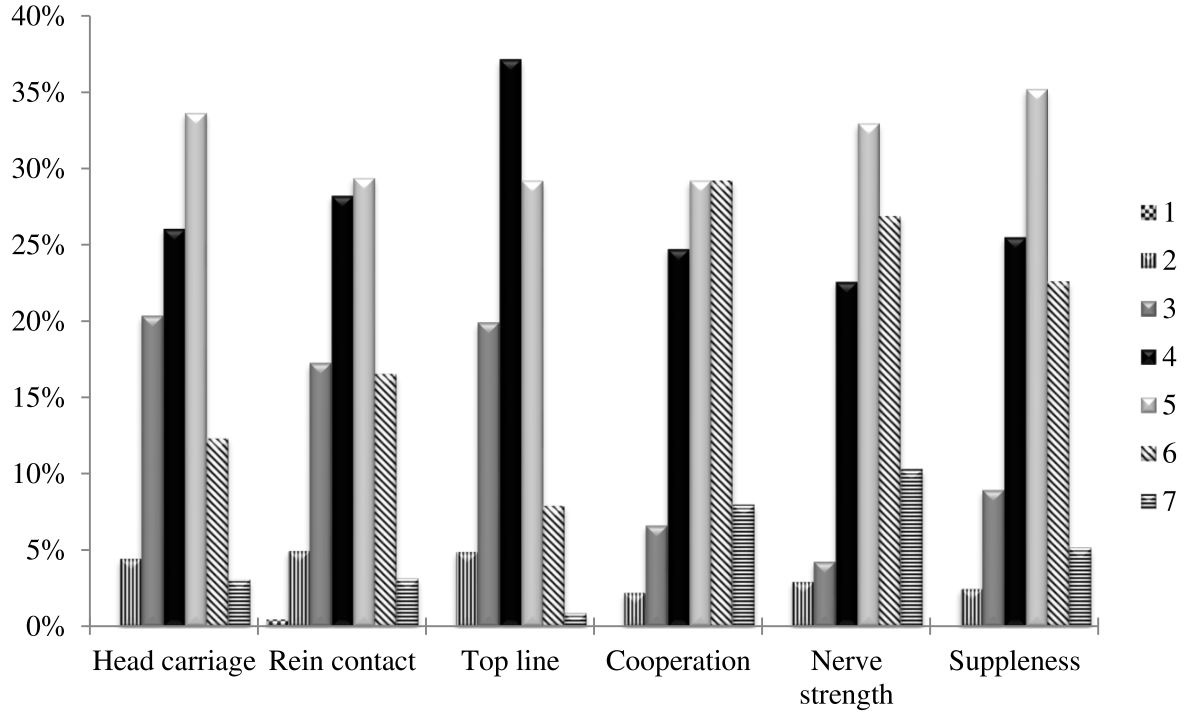

Figure 6. Distribution of scores for each of the six suppleness traits assessed by *judges* for horses in the sample group.

5.1.2 Principal component analysis

A principal component analyses (PCA) was carried out to identify possible underlying relationships between more than 2 variables. Scree test was used to determine the number of interpretable factors that could be extracted, and varimax rotation was used to identify empirical groupings of items that measured different traits. Table 8 shows the first 2 components for assessments of temperament traits made by both riders and judges. The total variance explained by the first 2 components was 78% for assessments made by the riders and 80% for assessments made by the judges. To interpret underlying components loadings of l>0.5 were considered. Using these criteria head carriage, rein contact, top line and overall suppleness loaded on the first component for the assessments made by riders, and spirit, cooperation, nerve strength and overall suppleness on the second component. For assessments made by judges; spirit, top line, cooperation, nerve strength and overall suppleness loaded on the first component and head carriage, rein contact, top line and overall suppleness loaded on the second component.

Table 8. Rotated factor pattern for temperament traits assessed by riders and judges for horses in the sample group

	Riders		**Judges**	
	Component 1	**Component 2**	**Component 1**	**Component 2**
Eigen value	4.79	0.68	4.67	0.95
Percentage variation	68%	9%	67%	14%
Spirit	0.25	0.80	0.88	0.14
Head carriage	0.79	0.29	0.25	0.82
Rein contact	0.78	0.35	0.19	0.91
Top line	0.85	0.36	0.53	0.66
Cooperation	0.49	0.77	0.84	0.37
Nerve strength	0.32	0.79	0.84	0.28
Suppleness	0.63	0.68	0.75	0.59

5.1.3 Genetic parameters

Heritabilities – all assessed horses

Estimated heritabilities, additive genetic variances, permanent environmental variances and residual variances for temperament traits assessed by judges for all horses assessed at breeding field tests in Iceland in 2014 and 2015 are presented in Table 9. Heritability estimates were low for all traits, with the highest being for rein contact and overall suppleness (0.19) and the lowest for head carriage (0.00). Average heritabilities for the temperament traits estimated in bivariate analyses are presented in Table 12. Heritability estimates did not differ much from univariate analyses, but the highest was for rein contact and overall suppleness (0.19) and the lowest for head carriage (0.04).

Table 9. Heritabilities (h^2), additive genetic (σ^2_a), permanent environmental (σ^2_{pe}) and residual (σ^2_e) variances (standard errors as subscripts) estimated in univariate analyses (using model 1) for temperament traits assessed by judges for all horses assessed at breeding field tests; heritabilities in bold indicate significant results

Trait	$\mathbf{h^2_{repeat}}$	$\boldsymbol{\sigma^2_{a\ repeat}}$	$\boldsymbol{\sigma^2_{pe\ repeat}}$	$\boldsymbol{\sigma^2_{e\ repeat}}$
Spirit	**0.14**	$0.02_{0.011}$	$0.05_{0.012}$	$0.10_{0.005}$
Head carriage	0.00	$0.00_{0.118}$	$0.11_{0.070}$	$1.22_{0.112}$
Rein contact	**0.19**	$0.25_{0.067}$	$0.03_{0.041}$	$1.07_{0.058}$
Top line	0.03	$0.04_{0.134}$	$0.37_{0.126}$	$0.84_{0.087}$
Cooperation	0.07	$0.10_{0.115}$	$0.07_{0.073}$	$1.18_{0.105}$
Nerve strength	**0.15**	$0.18_{0.064}$	$0.05_{0.046}$	$0.96_{0.052}$
Suppleness	**0.19**	$0.21_{0.054}$	$0.03_{0.032}$	$0.88_{0.047}$

Heritabilities – sample group

Heritabilities, additive genetic variances and residual variances estimated in univariate analyses for temperament traits assessed by judges and riders for the same horses at selected breeding field tests in Iceland in 2014 and 2015 are presented in Table 10, whereas results from bivariate analyses on heritability range are presented in Table 11. Heritabilities estimated for spirit assessed by riders were low and when assessed by judges they were low to moderate. Low heritabilities were estimated in univariate analyses for head carriage and top line when assessed by both riders and judges. In most of the bivariate analyses higher heritabilities were estimated however. Moderate heritabilities were estimated for head carriage assessed by riders when included with rein contact (0.21) and overall suppleness (0.35) and for top line when included with rein contact (0.25), nerve strength (0.31) and overall suppleness (0.30). Moderate to high heritabilities were estimated for head carriage assessed by judges when included with spirit (0.20) and top line (0.52) and for top line when included with spirit (0.72), head carriage (0.76) and nerve strength (0.22). Heritabilities estimated for rein contact were low to moderate when assessed by riders and moderate when assessed by judges, but varied on a similar range for both groups. Heritabilities estimated for cooperation, nerve strength and overall suppleness assessed by riders were moderate to high but low when assessed by the judges.

Table 10. Heritabilities (h^2), additive genetic (σ^2_a) and residual (σ^2_e) variances (standard errors as subscripts) estimated in univariate analyses for temperament traits assessed by *riders* and *judges* for horses in the sample group; heritabilities in bold indicate significant results

Trait	**h^2 riders**	**h^2 judges**	**σ^2 a riders**	**σ^2 a judges**	**σ^2 e riders**	**σ^2 e judges**
Spirit	0.08$_{0.098}$	0.15$_{0.123}$	0.03$_{0.037}$	0.03$_{0.023}$	0.34$_{0.042}$	0.16$_{0.023}$
Head carriage	0.02$_{0.195}$	0.01$_{0.204}$	0.03$_{0.298}$	0.01$_{0.264}$	1.50$_{0.323}$	1.29$_{0.286}$
Rein contact	0.22$_{0.123}$	0.21$_{0.129}$	0.34$_{0.199}$	0.29$_{0.184}$	1.22$_{0.193}$	1.09$_{0.178}$
Top line	0.07$_{0.193}$	0.00$_{0.248}$	0.15$_{0.396}$	0.00$_{0.258}$	1.89$_{0.420}$	1.04$_{0.270}$
Cooperation	0.31$_{0.247}$	0.00$_{0.193}$	0.64$_{0.538}$	0.00$_{0.248}$	1.46$_{0.504}$	1.29$_{0.257}$
Nerve strength	**0.39$_{0.122}$**	0.04$_{0.101}$	0.68$_{0.236}$	0.06$_{0.134}$	1.08$_{0.205}$	1.26$_{0.153}$
Suppleness	**0.45$_{0.134}$**	0.00$_{0.097}$	0.62$_{0.206}$	0.00$_{0.116}$	0.76$_{0.173}$	1.20$_{0.139}$

Table 11. Heritability (h^2) and standard error (S.E.) range estimated in bivariate analyses for temperament traits assessed by *riders* and *judges* for horses in the sample group; heritabilities in bold indicate significant results

Trait	**h^2 riders**	**S.E.**	**h^2 judges**	**S.E.**
Spirit	0.09 - 0.15	0.10 – 0.11	0.15 - 0.20	0.12
Head carriage	0.02 - 0.35	0.18 – 0.21	0.05 - **0.52**[*]	0.15 – 0.23
Rein contact	0.19 - **0.24**	0.11 – 0.12	0.21 - 0.23	0.13
Top line	0.04 - 0.31	0.16 – 0.20	0.07 – **0.76**[*]	0.18 – 0.27
Cooperation	0.23 - **0.49**	0.16 – 0.25	0.00 - 0.05	0.12 – 0.20
Nerve strength	**0.34 - 0.40**	0.12 – 0.13	0.04 - 0.12	0.10 – 0.12
Suppleness	**0.36 - 0.46**	0.13 – 0.14	0.03 - 0.07	0.11

* The highest estimated heritabilities for head carriage and top line were produced when these traits were analysed together

Correlations – all assessed horses

Genetic correlations estimated in bivariate analyses for temperament traits assessed by judges for all horses assessed at breeding field tests in Iceland in 2014 and 2015 are presented in Table 12. All genetic correlations were positive except correlation between rein contact and nerve strength (-0.26). All other correlations were estimated moderate to high.

Table 12. Genetic correlations (below the diagonal; standard errors as subscripts) and average heritability on diagonal estimated in bivariate analyses for temperament traits assessed by judges for all horses assessed at breeding field tests; parameters in bold indicate significant results

Trait	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Spirit (1)	**0.14**						
Head carriage (2)	**0.84$_{0.161}$**	0.04					
Rein contact (3)	**0.80$_{0.396}$**	1.00$_{0.968}$ [a]	**0.19**				
Top line (4)	**0.97$_{0.067}$**	**0.91$_{0.109}$** [a]	**1.00$_{0.221}$** [a]	0.06			
Cooperation (5)	**0.83$_{0.191}$**	**0.98$_{0.354}$** [a]	**1.00$_{0.491}$** [a]	**1.00$_{0.139}$** [c]	0.11		
Nerve strength (6)	**0.62$_{0.234}$**	**0.82$_{0.328}$** [a]	-0.26$_{0.771}$	1.00$_{0.587}$ [a]	**0.90$_{0.365}$**	**0.16**	
Suppleness (7)	**0.98$_{0.143}$**	**1.00$_{0.454}$** [b]	0.70$_{0.396}$	**1.00$_{0.130}$** [b]	**1.00$_{0.336}$** [a]	0.52$_{0.436}$ [a]	**0.19**

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)
[b] Convergence criteria used was lower (norm vector of $<10^{-5}$ instead of $<10^{-7}$)
[c] Convergence criteria used was lower (norm vector of $<10^{-4}$ instead of $<10^{-7}$)

Correlations – sample group

Estimated genetic correlations between temperament traits assessed by riders (below the diagonal) and judges (above the diagonal) and between assessments made by riders and judges (diagonal) are shown in Table 13. All correlations between temperament traits assessed by riders were positive, except correlation between head carriage and top line, and head carriage and cooperation, which had high standard error values and can be considered unreliable. All other genetic correlations were moderately high to high. All correlations between temperament traits assessed by judges were positive, except genetic correlation between cooperation and overall suppleness, which had high standard error values. High standard error values were estimated for all correlations between cooperation and other traits, and most of the correlations estimated between head carriage and other traits. All genetic correlations were moderate to high, where the correlation between spirit and rein contact (0.53) and spirit and nerve strength (0.50) were the lowest. All genetic correlations between assessments made by riders and judges were positive and ranged from moderate to high. The lowest estimated correlation was between assessments for rein contact (0.45) and for spirit (0.57).

Table 13. Genetic correlations estimated in bivariate analyses between temperament traits assessed by *riders* (below the diagonal), *judges* (above the diagonal) and between assessments made by *riders and judges* (diagonal) for horses in the sample group; correlations in bold indicate significant results

Trait	1	2	3	4	5	6	7
Spirit (1)	$0.57_{0.494}$	$1.00_{0.581}^{b}$	$0.53_{0.423}$	$\mathbf{0.99_{0.189}}$	$1.00_{1.003}^{b}$	$0.50_{0.716}$	$1.00_{0.570}^{b}$
Head carriage (2)	$1.00_{0.556}^{b}$	$\mathbf{1.00_{0.178}^{b}}$	$1.00_{1.173}^{b}$	$\mathbf{1.00_{0.130}}^{c}$	$1.00_{69.618}^{a}$	$1.00_{1.384}^{b}$	$1.00_{1.351}^{c}$
Rein contact (3)	$\mathbf{1.00_{0.389}}^{b}$	$\mathbf{1.00_{0.342}}^{b}$	$0.45_{0.370}$	$1.00_{0.759}^{b}$	$1.00_{40.439}^{a}$	$0.92_{0.510}$	$\mathbf{1.00_{0.348}}^{b}$
Top line (4)	$1.00_{0.602}^{c}$	$-0.73_{7.430}$	$\mathbf{1.00_{0.241}}^{b}$	$\mathbf{1.00_{0.203}^{c}}$	$1.00_{385.727}^{c}$	$1.00_{0.513}^{b}$	$\mathbf{1.00_{0.477}}^{a}$
Cooperation (5)	$\mathbf{0.97_{0.246}}$	$-1.00_{6.248}^{b}$	$\mathbf{0.74_{0.197}}$	$\mathbf{0.86_{0.342}}$	$1.00_{1.685}^{b}$	$1.00_{5.924}^{a}$	$-1.00_{19.935}^{a}$
Nerve strength (6)	$\mathbf{0.80_{0.337}}$	$\mathbf{1.00_{0.389}}^{b}$	$\mathbf{1.00_{0.225}}^{c}$	$\mathbf{1.00_{0.230}}^{c}$	$\mathbf{0.86_{0.164}}$	$0.82_{0.710}$	$1.00_{0.511}^{c}$
Suppleness (7)	$\mathbf{1.00_{0.223}}^{b}$	$\mathbf{0.96_{0.163}}$	$\mathbf{1.00_{0.116}}^{c}$	$\mathbf{0.95_{0.138}}$	$\mathbf{0.90_{0.096}}$	$\mathbf{0.97_{0.081}}$	$1.00_{0.519}^{c}$

[b] Convergence criteria used was lower (norm vector of $<10^{-5}$ instead of $<10^{-7}$)
[c] Convergence criteria used was lower (norm vector of $<10^{-4}$ instead of $<10^{-7}$)

Estimated phenotypic correlations between temperament traits assessed by riders (below the diagonal) and judges (above the diagonal) and between assessments made by riders and judges (diagonal) are shown in Table 14. All correlations were positive. All correlations between temperament traits assessed by riders were moderate to high, where the correlation between cooperation and overall suppleness (0.83) was the highest and the correlation between rein contact and nerve strength (0.48) was the lowest. All correlations between temperament traits assessed by judges were moderate to high, where the correlation between cooperation and overall suppleness (0.87) was the highest and the correlation between spirit and rein contact (0.35) was the lowest. Phenotypic correlations between assessments made by riders and judges ranged from low to moderate, where the correlation for assessments made for spirit (0.61) was the highest and the correlations for assessments made for rein contact (0.22) and nerve strength (0.24) were the lowest.

Table 14. Phenotypic correlations estimated in bivariate analyses for temperament traits assessed by *riders* (below the diagonal), *judges* (above the diagonal) and between assessments made by *riders and judges* (diagonal) for horses in the sample group

Trait	1	2	3	4	5	6	7
Spirit (1)	0.61	0.43[b]	0.35	0.59	0.72[b]	0.51	0.58[b]
Head carriage (2)	0.51[b]	0.45[b]	0.66[b]	0.59[c]	0.48[a]	0.42[b]	0.64[c]
Rein contact (3)	0.50[b]	0.58[b]	0.22	0.67[b]	0.51[a]	0.55	0.77[b]
Top line (4)	0.49[c]	0.73	0.71[b]	0.40[c]	0.65[c]	0.58[b]	0.74[a]
Cooperation (5)	0.63	0.58[b]	0.66	0.71	0.38[b]	0.77[a]	0.87[a]
Nerve strength (6)	0.49	0.49[b]	0.48[c]	0.55[c]	0.69	0.24	0.83[c]
Suppleness (7)	0.64[b]	0.68	0.74[c]	0.76	0.83	0.70	0.38[c]

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)
[b] Convergence criteria used was lower (norm vector of $<10^{-5}$ instead of $<10^{-7}$)
[c] Convergence criteria used was lower (norm vector of $<10^{-4}$ instead of $<10^{-7}$)

5.2 Part II

5.2.1 Descriptive statistics

Seventeen general temperament traits were assessed in fifteen questions included in the survey about general temperament traits. The first two questions referred to time in training and training level when the horse was subjected to a breeding field test, and were intended to give information about how fast the horse developed as a response to training. Statistics for scores given for general temperament traits in the survey are presented in Table 15. Mean score for time in training was 2.00 (4 point scale) and for training level 2.29 (3 point scale). Mean scores for other traits scored on a 4 point scale ranged between 2.90 and 3.50. The highest mean scores were for behaviour towards humans (towards the score for very calm behaviour) and behaviour while being ridden away from stable (towards the score for very willing behaviour). The lowest mean score was for reaction to new aids while being broken in (towards the score for relaxed reaction). The scores ranged from 1 to 4 for all these traits except the trait behaviour predictability, no horse was scored as very unpredictable. Distribution for most traits was fairly symmetric except the traits behaviour towards humans and behaviour while being ridden away from stable which had strong peaks skewed to the left. Time in training was the only trait with an upward skewness. Nevertheless, all traits tested significant for normal distribution according to Anderson-Darling test ($p<0.005$).

Table 15. Number of assessments, mean score ($\bar{x}$), standard deviation (S.D.), range, skewness and kurtosis for assessments made by owners/breeders/trainers in the survey

Question	No. of assessments	$\bar{x}$	S.D.	Range*	Skewness	Kurtosis
Q1 Time in training	333	2.00	0.94	1 - 4	0.70	-0.38
Q2 Training level	329	2.29	0.69	1 - 3	-0.46	-0.86
Q3 Reaction to a novel object	338	3.14	0.75	1 - 4	-0.53	-0.19
Q4 Action when seeing a novel object	316	3.20	0.70	1 - 4	-0.75	0.92
Q5 Reaction to novel/loud sound	336	3.10	0.72	1 - 4	-0.53	0.19
Q6 Reaction to temporary isolation	337	3.21	0.65	1 - 4	-0.36	-0.12
Q7 Reaction to new environment	337	3.09	0.66	1 - 4	-0.53	0.84
Q8 Behaviour towards humans	339	3.50	0.65	1 - 4	-1.20	1.39
Q9 Behaviour towards other horses	334	3.13	0.66	1 - 4	-0.27	-0.20
Q10 Reaction to disturbance while being trained	336	3.09	0.76	1 - 4	-0.48	-0.23
Q11a Reaction to new aids while being broken in (apt/unapt)	325	3.18	0.66	1 - 4	-0.40	0.11
Q11b Reaction to new aids while being broken in (tension)	324	2.90	0.72	1 - 4	-0.31	0.00
Q12a Behaviour while being trained (cooperation)	337	3.31	0.62	1 - 4	-0.39	-0.24
Q12b Behaviour while being trained (tension)	337	2.97	0.64	1 - 4	-0.18	0.09
Q13 Behaviour while being ridden away from stable	336	3.50	0.62	1 - 4	-1.16	1.72
Q14 General cooperation in handling	338	3.45	0.62	1 - 4	-0.89	0.87
Q15 Behaviour predictability	339	3.32	0.57	2 - 4	-0.14	-0.62

*Maximum score for all questions is 4 and minimum score is 1, except for question about training level where maximum score is 3 and minimum score is 1.

Respondents to the survey were asked how well the score for spirit received at a breeding field test conforms to the general temperament of the horse being assessed. 69% of respondents answered well or very well and 31% answered badly or very badly. Of these 31%, 86% answered it was underestimated and 14% overestimated. Only one respondent did not answer this question. Distribution of answers to this question is presented in Figure 7.

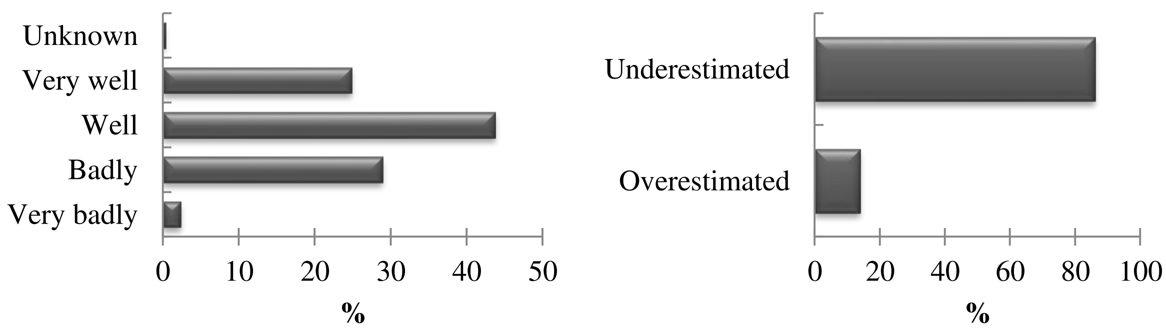

Figure 7. Distribution of answers to how well score for Spirit conforms to the general temperament of the horse (to the left) and how the inconsistency is explained if it conforms badly or very badly to the score (to the right), answered by owners/breeders/trainers in the survey.

Figure 8 shows the ranking of specific temperament traits according to importance by respondents of the survey, where 1 is the most important trait and 7 is the least important trait. The trait most frequently ranked as the most important by the respondents was suppleness with the highest frequency of number 1 ranking, and the least important trait was fiery with the highest frequency of number 6 ranking. The second most important trait was nerve strength. Few respondents used the option "Other" to explain that all traits were internally related and equally important and could therefore not be ranked according to importance.

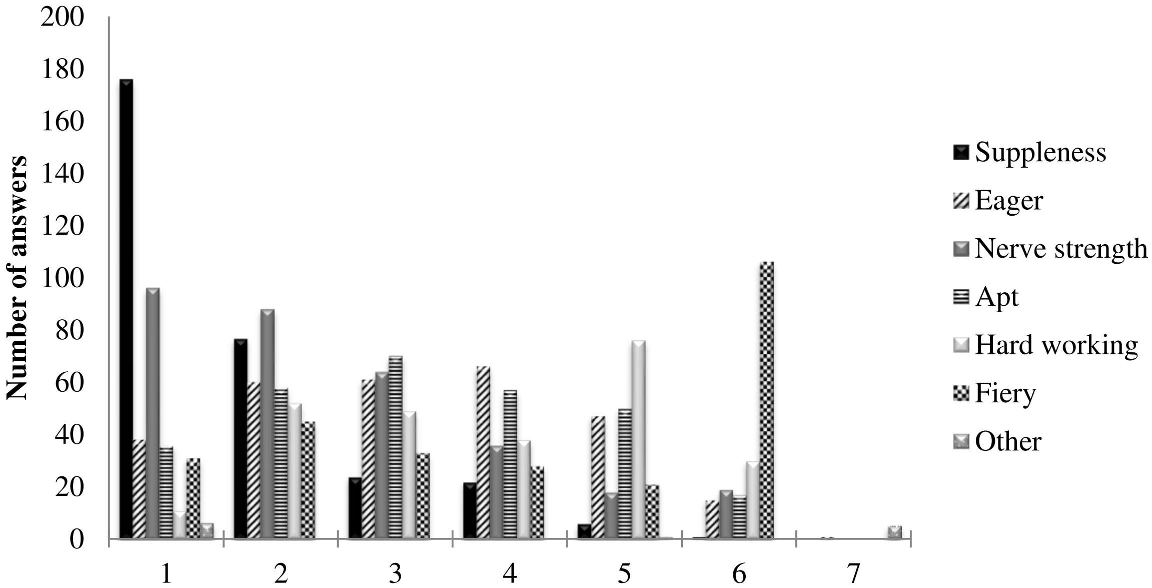

Figure 8. Ranking of traits according to importance (1 being the most important, 7 being the least important) by owners/breeders/trainers in the survey.

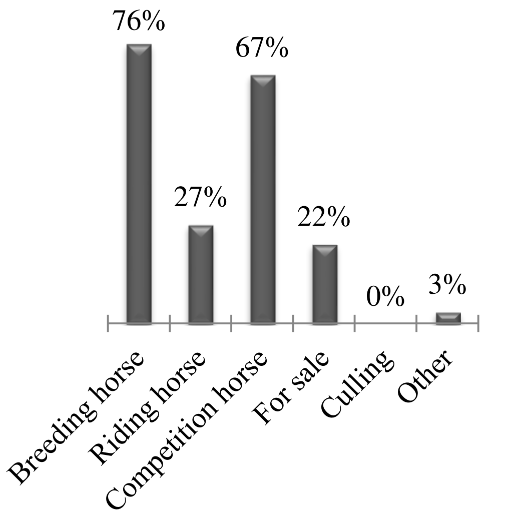

Figure 9. Distribution of answers to what the future role of the horse is expected to be, answered by owners/breeders/trainers in the survey.

The most common intended future roles of the horses assessed by the respondents in the survey were breeding and competition. Multiple choices were allowed in this question and 50% of the horses were intended to be used for both breeding and competition and 20% for both breeding and riding. The future role was undecided for a few horses and some horses had already been sold. Figure 9 shows the distribution of intended future roles of the horses assessed in the survey.

Statistics for scores given for temperament traits assessed by judges at breeding field tests for the horses also assessed in the survey are presented in Table 16. All horses assessed in the survey had received score for spirit where the mean score was 8.51 with scores ranging from 7.5 to 9.5. Assessments for suppleness traits were missing for 22 horses assessed at breeding field tests in 2014 and 18 horses assessed in 2015. Scores for suppleness traits ranged from 2 to 7 except for rein contact, but one horse had received the score 1 for this trait. The mean score for these traits ranged from 4.73 to 5.29, where score for cooperation was the highest and score for top line was the lowest. Distribution of scores was fairly symmetric and all traits tested significant for normal distribution according to Anderson-Darling test ($p<0.005$).

Table 16. Number of assessments, mean score ($\bar{x}$), standard deviation (S.D.), range, skewness and kurtosis for temperament traits assessed by judges at breeding field tests, for the horses that were assessed by their owners/breeders/trainers in the survey

Trait	No. of assessments	$\bar{x}$	S.D.	Range*	Skewness	Kurtosis
Spirit	339	8.51	0.42	7.5 -9.5	0.22	-0.29
Head carriage	129	4.91	1.23	2 - 7	-0.32	-0.15
Rein contact	299	4.79	1.18	1 - 7	-0.29	-0.13
Top line	129	4.73	1.17	2 - 7	-0.17	-0.36
Cooperation	129	5.29	1.20	2 - 7	-0.42	-0.23
Nerve strength	299	5.22	1.18	2 - 7	-0.46	-0.05
Suppleness total	299	5.11	1.06	2 - 7	-0.26	-0.14

*Maximum score on scale for spirit is 10.0 and minimum score is 5.0. Maximum score on scale for head carriage, rein contact, top line, cooperation, nerve strength and suppleness total is 7.0 and minimum score is 1.0.

EBVs for spirit differed between horses that received the highest and lowest score for nerve strength, but the mean EBV for horses that received the highest score was 116 and 107 for the horses that received the lowest score.

5.2.2 Principal component analysis

A principal component analyses (PCA) was carried out to identify possible underlying relationships between questions in the survey. Table 17 shows the first 4 principal components of assessments for general temperament traits made by owners/trainers/breeders. The total variance explained by the first 4 components was 61% (eigenvalues$\geq$1). To interpret underlying components, loadings of $l>0.4$ were considered. Using these criteria, questions Q3, Q4, Q5, Q7, Q8, Q10, Q11b, Q12b and Q15 were found to load on the first component, which was subsequently labelled the general nerve strength component. Questions Q11a, Q12a, Q13 and Q14 loaded on the second component which was labelled the general cooperation component. Questions Q6, Q7, Q9 and Q13 loaded on the third component which was labelled the independence/sociability component and questions Q1 and Q2 loaded on the fourth component which was labelled the training response component.

Table 17. Rotated factor pattern for questions answered by owners/breeders/trainers in the survey

	Components			
	1 General nerve strength	**2 General cooperation**	**3 Independence / sociability**	**4 Training response**
Eigen value	5.80	1.97	1.47	1.12
Percentage variation	34%	12%	9%	7%
Q1 Time in training	0.03	-0.08	-0.09	0.85
Q2 Training level	0.10	0.13	0.09	0.83
Q3 Reaction to a novel object	0.83	-0.01	0.19	-0.02
Q4 Action when seeing a novel object	0.72	0.02	0.13	-0.09
Q5 Reaction to novel/loud sound	0.79	0.00	0.23	0.09
Q6 Reaction to temporary isolation	0.35	-0.08	0.69	0.05
Q7 Reaction to new environment	0.52	0.06	0.55	0.07
Q8 Behaviour towards humans	0.66	0.12	0.21	0.08
Q9 Behaviour towards other horses	0.09	0.16	0.62	-0.11
Q10 Reaction to disturbance while being trained	0.81	0.01	0.22	0.05
Q11a Reaction to new aids while being broken in (apt/unapt)	-0.16	0.72	0.15	-0.10
Q11b Reaction to new aids while being broken in (tension)	0.77	0.08	-0.03	0.06
Q12a Behaviour while being trained (cooperation)	0.34	0.75	-0.08	0.02
Q12b Behaviour while being trained (tension)	0.76	0.10	0.05	0.13
Q13 Behaviour while being ridden away from stable	0.02	0.48	0.57	0.10
Q14 General cooperation in handling	0.19	0.75	0.15	0.12
Q15 Behaviour predictability	0.61	0.27	0.05	0.01

5.2.3 Genetic parameters

Heritabilities

Estimated heritabilities, additive genetic variances and residual variances for general temperament traits assessed by owners, breeders and/or trainers in the survey for horses tested at breeding field tests in Iceland in 2014 and 2015 are presented in Table 18. Heritabilities were estimated in a bivariate analysis using model 4 with the interaction between age and sex, and training level as fixed effects, except for the trait training level which was estimated using model 5 with age and sex, and training time as fixed effects. Heritabilities ranged from low to high, and were estimated with high standard errors in this small data set. The general nerve strength component included traits with estimated heritabilities ranging from low to moderate, where the lowest was for behaviour predictability (0.01) and the highest was for action when seeing a novel object (0.27). Heritabilities estimated for traits within the general cooperation component were low, where the lowest was for behaviour while being ridden away from stable and general cooperation in handling (zero heritability estimates) and the highest was for behaviour while being trained (cooperation) (0.08). Heritabilities estimated for traits within the independence/sociability component were low, where the lowest was for reaction to temporary isolation and behaviour while being ridden away from stable (zero heritability estimates) and the highest being for behaviour towards other horses (0.18). Heritability estimated for time in training was low (0.16) and high for training level (0.46). Heritabilities estimated in bivariate analyses using model 3 for general temperament traits and the trait spirit are presented in Table 1-4 in Appendix VI.

Table 18. Heritabilities (h^2), additive genetic (σ^2_a) and residual (σ^2_e) variances (standard errors as subscripts) estimated in bivariate analyses with the trait spirit (using model 4) for traits assessed by owners/breeders/trainers in the survey, and genetic correlation (r_g) (standard errors as subscripts) and phenotypic correlation (r_p) between spirit and the same traits; heritabilities in bold indicate significant results

	h^2	σ^2_a	σ^2_e	r_g	r_p
Spirit	$0.34_{0.185}$	$0.05_{0.030}$	$0.10_{0.028}$		
General nerve strength component					
Q3 Reaction to a novel object	$0.09_{0.139}$	$0.05_{0.078}$	$0.51_{0.084}$	$-0.17_{0.665}$	-0.12
Q4 Action when seeing a novel object	$0.27_{0.191}$	$0.13_{0.096}$	$0.36_{0.091}$	$-0.74_{0.439}$	-0.11
Q5 Reaction to novel/loud sound	$0.24_{0.160}$	$0.13_{0.085}$	$0.39_{0.082}$	$-0.84_{0.301}$	-0.07
Q7 Reaction to new environment*	$0.07_{0.139}$	$0.03_{0.062}$	$0.41_{0.067}$	$-0.41_{0.758}$	-0.02
Q8 Behaviour towards humans	$0.10_{0.133}$	$0.04_{0.053}$	$0.36_{0.057}$	$-0.74_{0.522}$	-0.07
Q10 Reaction to disturbance while being trained	$0.19_{0.172}$	$0.11_{0.096}$	$0.44_{0.094}$	$-0.73_{0.386}$	-0.11
Q11b Reaction to new aids while being broken in (tension)	$0.26_{0.186}$	$0.13_{0.098}$	$0.38_{0.093}$	$-0.97_{0.305}$	-0.10
Q12b Behaviour while being trained (tension)	$0.09_{0.141}$	$0.04_{0.055}$	$0.35_{0.058}$	$-0.81_{0.647}$	-0.01
Q15 Behaviour predictability	$0.01_{0.165}$	$0.00_{0.054}$	$0.32_{0.056}$	$-1.00_{6.842}$ [a]	0.04 [a]
General cooperation component					
Q11a Reaction to new aids while being broken in (apt/unapt)	$0.04_{0.158}$	$0.02_{0.068}$	$0.41_{0.072}$	$1.00_{2.383}$ [a]	0.05 [a]
Q12a Behaviour while being trained (cooperation)	$0.08_{0.151}$	$0.03_{0.057}$	$0.35_{0.061}$	$0.57_{0.817}$	0.13
Q13 Behaviour while being ridden away from stable*	$0.00_{0.135}$	$0.00_{0.050}$	$0.37_{0.056}$	$1.00_{14.01}$ [a]	0.02 [a]
Q14 General cooperation in handling	$0.00_{0.128}$	$0.00_{0.048}$	$0.37_{0.054}$	$-1.00_{13.37}$ [a]	0.08 [a]
Independence/sociability component					
Q6 Reaction to temporary isolation	$0.00_{0.129}$	$0.00_{0.053}$	$0.41_{0.059}$	$-1.00_{41.01}$	-0.03
Q9 Behaviour towards other horses	$0.18_{0.167}$	$0.08_{0.071}$	$0.34_{0.071}$	$0.01_{0.523}$	0.03
Training response component					
Q1 Time in training	$0.16_{0.184}$	$0.07_{0.078}$	$0.35_{0.077}$	$-0.60_{0.609}$	0.07
Q2 Training level**	$\mathbf{0.46_{0.188}}$	$0.17_{0.074}$	$0.19_{0.065}$	$0.42_{0.301}$	0.25

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)

* Reaction to new environment and behaviour while being ridden away from stable are also included in the independence/sociability component.
** Genetic parameters for training level were estimated using model 5.

Correlations between spirit and general temperament traits

Genetic and phenotypic correlations between the trait spirit and the general temperament traits estimated in bivariate analyses using model 4 are presented in Table 18. All genetic correlations between spirit and general temperament traits within the general nerve strength component were negative and ranged from low to high, where the highest was between spirit and reaction to novel/loud sound (-0.84) and the lowest between spirit and reaction to a novel object (-0.17). Most of the genetic correlations between spirit and traits within the general cooperation component were estimated with high standard error values. Behaviour while being trained (cooperation) was estimated to be moderately and positively correlated to spirit (0.57). Most of the genetic correlations between spirit and traits within the independence/sociability component were estimated with high standard error values. Genetic correlation between time in training and spirit was estimated moderate and negative (-0.60) but moderate and positive between training level and spirit (0.42). Estimated phenotypic correlations between spirit and all the general temperament traits were low where the highest was between spirit and training level (0.25). Genetic and phenotypic correlations between the general temperament traits and the trait spirit estimated in bivariate analyses using model 3 are presented in Table 1 in Appendix VI.

Correlations between temperament traits assessed at breeding field tests and general temperament traits

Phenotypic correlations between temperament traits assessed by judges at breeding field tests and general temperament traits assessed by owners, breeders and/or trainers in the survey were estimated in a multivariate analyses using Spearman rank-order correlation. The results are presented in Table 19. Estimated correlations were low (0.01-0.36) with the highest being between training level and spirit. Of the temperament traits assessed by judges, nerve strength had the highest correlation on average to the general temperament traits assessed by owners, breeders and/or trainers although the correlation was weak (0.12-0.21).

Table 19. Phenotypic correlation estimated in a multivariate analysis (Spearman) for traits assessed by owners/breeders/trainers in the survey and temperament traits assessed by judges at breeding field tests; correlations in bold indicate significant results ($p<0.05$)

	Spirit	**Head carriage**	**Rein contact**	**Top line**	**Coope ration**	**Nerve strength**	**Supple ness**
General nerve strength component							
Q3 Reaction to a novel object	-0.06	-0.09	-0.08	0.04	-0.01	0.11	-0.02
Q4 Action when seeing a novel object	-0.07	-0.07	**-0.13**	-0.03	0.05	0.04	-0.08
Q5 Reaction to novel/loud sound	0.01	0.03	-0.05	0.12	0.09	**0.20**	0.07
Q7 Reaction to new environment*	0.02	0.06	-0.11	0.01	0.04	0.04	-0.05
Q8 Behaviour towards humans	0.03	-0.04	**-0.14**	0.03	0.01	0.07	-0.04
Q10 Reaction to disturbance while being trained	-0.01	0.09	-0.04	0.05	0.05	**0.14**	0.04
Q11b Reaction to new aids while being broken in (tension)	-0.03	0.04	-0.01	0.08	0.02	**0.19**	0.07
Q12b Behaviour while being trained (tension)	0.08	0.04	-0.02	0.08	0.02	**0.20**	0.09
Q15 Behaviour predictability	0.06	0.14	0.01	**0.20**	0.03	0.04	0.04
General cooperation component							
Q11a Reaction to new aids while being broken in (apt/unapt)	0.03	0.11	0.04	-0.04	0.06	0.01	0.04
Q12a Behaviour while being trained (cooperation)	**0.13**	0.14	0.10	**0.20**	**0.19**	**0.15**	**0.17**
Q13 Behaviour while being ridden away from stable*	0.06	-0.01	**-0.12**	-0.08	0.10	-0.01	0.00
Q14 General cooperation in handling	**0.14**	0.07	0.00	**0.20**	0.11	**0.13**	0.09
Independence/sociability component							
Q6 Reaction to temporary isolation	0.04	0.12	0.01	0.05	-0.04	0.09	0.07
Q9 Behaviour towards other horses	0.10	0.06	0.03	0.06	0.09	**0.19**	0.10
Training response component							
Q1 Time in training	**0.22**	0.17	0.01	0.14	0.16	**0.12**	0.07
Q2 Training level	**0.36**	**0.26**	0.05	**0.26**	**0.25**	**0.18**	**0.14**

* Reaction to new environment and behaviour while being ridden away from stable are also included in the independence/sociability component.

Correlations between general temperament traits

Genetic and phenotypic correlations estimated in bivariate analyses for general temperament traits, using model 3, are presented in Table 2-4 in Appendix VI.

Genetic correlations estimated between traits loading on the general nerve strength component ranged from low to high, but several traits were estimated with high standard error values. Reaction to disturbance while being trained was estimated with high genetic correlation to most of the other traits. Reaction to a novel object was estimated with low genetic correlation to most of the other traits. All estimated phenotypic correlations were positive and ranged from moderate to moderately high (0.35-0.68), where the correlation between reaction to novel/loud sound and reaction to disturbance while being trained was the highest.

Genetic correlations between traits loading on the general cooperation component were estimated with high standard error values and were considered unreliable results. Estimated phenotypic correlations were moderate and positive, where the highest was between behaviour while being trained (cooperation) and general cooperation in handling (0.52) and the lowest between behaviour while being trained (cooperation) and behaviour while being ridden away from stable (0.27).

Genetic correlations for traits loading on the independence/sociability component were estimated with high standard error values and were considered unreliable results. Phenotypic correlations were all positive and ranged from low to moderate, where the highest was between reaction to temporary isolation and reaction to new environment (0.44) and the lowest between reaction to temporary isolation and behaviour while being ridden away from stable (0.15).

The genetic correlation between time in training and training level was 0.58±0.31 and the phenotypic correlation was 0.31.

5.3 Part III

5.3.1 Descriptive statistics

Figure 10 shows the distribution of answers to what the reason for culling was, answered by owners of horses culled during the period of September 2014 to January 2015. Some owners gave several reasons for culling. The most common reason was culling due to temperament faults such as nervousness, tension, stubbornness, laziness, lack of cooperation and horses displaying out of control behaviour and frequent attempts to bolt or rear and/or buck. High age and the horse having no purpose or role were also frequently given reasons for culling. Traits included in the breeding goal of Icelandic horses such as temperament, riding ability, conformation, health and fertility compiled to 48% of culled horses, as a reason for culling. Among different reasons for culling; accident, temperament fault and high age were most commonly given as a single reason but conformation fault, meat production and fertility problems were most commonly given as a reason combined with another reason. The most common combination with temperament fault was lack of ride ability and no role or purpose.

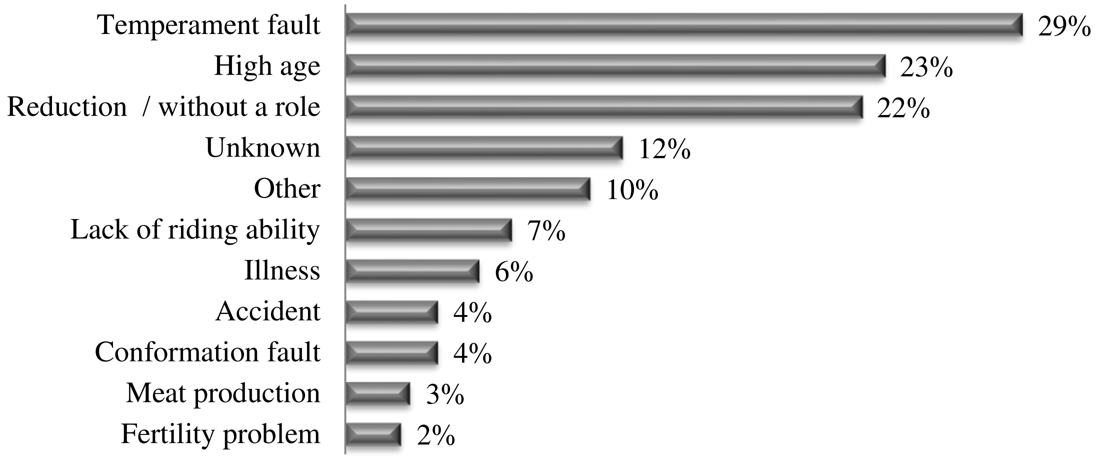

Figure 10. Distribution of horses culled due to various reasons during the period of September 2014 to January 2015.

Figure 11 shows the distribution of past roles of the horses that were culled during the period of September 2014 and January 2015. Most of the horses served as a riding/leisure horse during their lifetime. Young horses that were withdrawn from further training for different reasons, and horses that never had any role or purpose accounted for 40% of the sample. Some horses had several roles and the most common combination was a riding horse and a breeding horse. Of all the breeding horses that were culled, 29% had been subjected to a breeding field test in the past, which represent 4% of the sample. The average age of the breeding horses was 19 years at the time of culling.

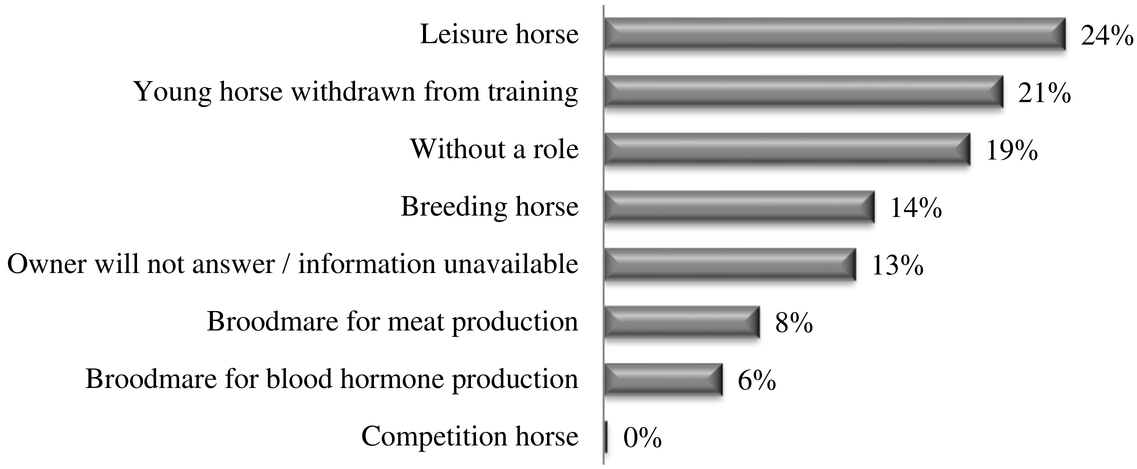

Figure 11. Distribution of past roles of the horses that were culled during the period of September 2014 to January 2015.

Of all the horses culled due to temperament faults, 55% were young horses that were withdrawn from further training, 21% were riding/leisure horses and 20% were horses that had no role or purpose. Of those horses that were culled due to high age, 47% were riding/leisure horses and 44% were breeding horses where the average age of both groups was

21 years at the time of culling. 61% of the horses that were culled because of lack of riding ability were young horses that were withdrawn from further training. Details of past roles of horses culled due to various reasons are presented in Table 20.

Table 20. Past roles of horses that were culled due to various reasons during the period of September 2014 to January 2015 (table should be read columnwise)

	Reason for culling						
Role*	Tempera-ment fault	High age	Reduction / without a role	Lack of riding ability	Illness	Confor-mation fault	Fertility problem
Leisure horse	21%	47%	16%	33%	48%	15%	0%
Young horse withdrawn from training	55%	0%	9%	61%	10%	40%	0%
Without a role	20%	3%	47%	14%	14%	40%	8%
Breeding horse	3%	44%	20%	0%	24%	10%	67%
Broodmare for meat production	1%	12%	15%	0%	14%	10%	17%
Broodmare for blood hormone production	7%	4%	0%	0%	3%	0%	17%

* Several horses had more than one role or purpose during their lifetime.

Of all the horses culled due to temperament faults, 57.1% were geldings/stallions and 42.9% were mares. The most common age of culled horses ranged from 3 years old to 7 years old, where 5 years old was the most common age. Figure 12 shows the age distribution of horses culled due to temperament faults.

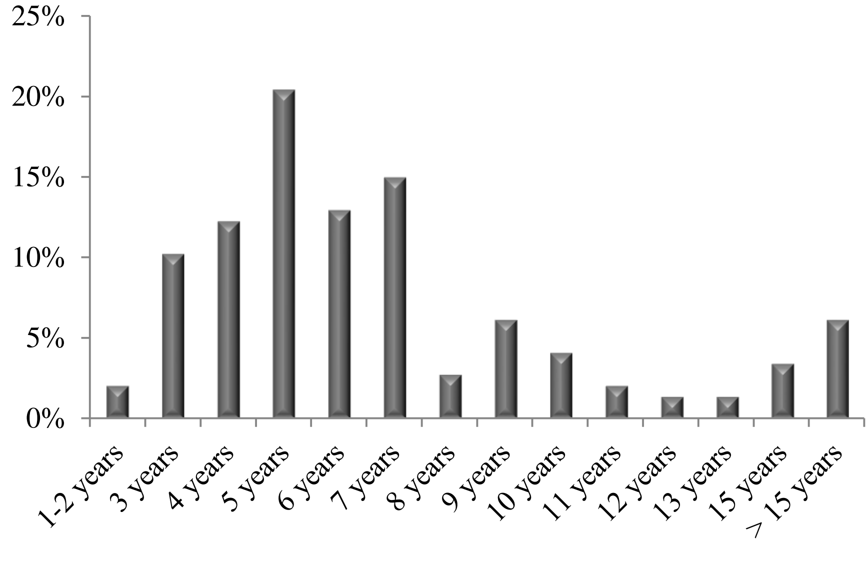

Figure 12. Age distribution of horses culled due to temperament faults during the period of September 2014 and January 2015.

Mean EBVs for spirit, riding ability and total score for conformation and riding ability for young horses that were withdrawn from further training and culled for different reasons are presented in Table 21. Breeding values were estimated by the FAI in the autumn of 2015. Mean EBVs were similar for all traits and did not vary significantly between groups culled for different reasons.

Table 21. Number of individuals (N), mean estimated breeding values (EBV) and range of EBV for spirit, riding ability and total score for conformation and riding ability for young horses that were withdrawn from further training and culled for different reasons

Reason for culling	N	**Spirit**		**Riding ability**		**Total score**	
		Mean EBV	Range	Mean EBV	Range	Mean EBV	Range
Temperament fault	80	103	86-117	102	87-117	103	86-119
Lack of riding ability	22	102	93-114	102	93-115	102	94-115
Without a role/purpose	10	104	98-109	104	97-111	105	97-112
Conformation fault	8	102	87-110	102	84-111	103	82-114

Mean accuracy of EBVs ranged between 40%-68%

6 Discussion

The main focus of the thesis was to evaluate the current assessment method for the trait spirit in breeding field tests for Icelandic horses. As mentioned before, the assessment method has been a topic of discussion within the Icelandic horse society for several years, where many consider that the current method favours nervous and tense horses, but supple and stable horses better suited for general riders receive lower scores for spirit. It has also been suggested that these supposedly tense horses receiving higher scores for spirit often give nervous offspring. Practical experience of people within the horse society and their general impression of the breeding progress regarding the temperament of the Icelandic horses, as was expressed at the symposium in 2013, must not be ignored. This is especially important as temperament traits have been ranked as more important than any other performance trait by a majority of horse enthusiast within different disciplines according to several studies (Gille & Spiller, 2010; Gille *et al.*, 2010; Górecka-Bruzda *et al.*, 2011b; Graf *et al.*, 2013). Therefore it seems imperative to evaluate the assessment method and strive for improvements if needed, especially considering the impact on human safety and economic aspects.

The first part of the study aimed to estimate genetic parameters for temperament traits assessed by breeding judges at breeding field tests and to investigate whether these traits possess genetic potential to improve the current assessment method for spirit. The first part was also to investigate the consistency between assessments made by breeding judges and BFT riders for the temperament traits in order to examine if there are any factors within the traits that the judges are less able to comprehend in their assessments than the riders. The second part aimed to investigate how well the score for spirit describes the general temperament of horses. Genetic parameters were estimated for general temperament traits assessed in the everyday environment by owners, breeders and/or trainers of horses that had previously been assessed at breeding field tests. This was done to examine if there is a relationship between the assessments made for temperament traits at breeding field tests and general temperament traits observed at home. The samples in both parts seemed to be fairly representative of all the horses presented at breeding field tests in Iceland in 2014 and 2015, based on distribution of scores and mean EBVs. Results from both parts were logical and gave fairly clear indications of genetic variances in the traits assessed. However, parameters were estimated with high standard errors in both samples due to the small number of individuals included in the datasets. This limits the possibility to draw firm conclusions without further investigation. The third part was to estimate the frequency of culling due to temperament faults in order to gain information on temperament in horses that are usually not included in the preselected group of horses that are presented at breeding field tests. Thereby to investigate the possibility of integrating such information in the breeding evaluation for spirit, as means to increase the accuracy of the evaluation by including a higher proportion of the Icelandic horse population. Age distribution in the data collected for this part was similar to that for all the horses culled during the period investigated, indicating that the studied data sample was representative, but further investigation using a larger dataset representing a longer time period is suggested.

The study approaches different aspects of the Icelandic horse population by including breeding horses preselected to be assessed at breeding field tests and culled horses that have never been assessed, with a small overlap between these two groups. The study touches upon the preselection of the horses that are considered to have the most desired temperament and the ones that are considered to have undesired temperament. This was partly reflected in the mean EBVs for total score and spirit in the different groups in the study, where the mean EBV of culled horses was lower than the mean of the reference population and the mean EBV of horses shown at breeding field tests was higher than the mean. This is nevertheless logical, considering the difference in selection criteria for horses that are culled and horses that are presented at breeding field tests. Also, the breeding values estimated for horses that are not presented at breeding field tests are usually based on little information and are estimated with low accuracy as is evident in the sample of culled horses. It would therefore be of advantage for the breeding evaluation to obtain more information on these horses. The study also approaches different groups of people within the horse society, giving the opportunity to gain insight to the different requirements of each group and adjust the breeding system accordingly to breed horses suited for both general and professional riders.

6.1 Temperament traits assessed at breeding field tests

Most of the temperament traits assessed at breeding field tests had considerable genetic variance and strong and positive internal genetic correlations, which indicates a certain reliability of the traits and justification of the assessments. A relatively low heritability for spirit was estimated based on the sample of horses assessed by both riders and judges, compared to what has been reported before (Árnason & Sigurðsson, 2004; Albertsdóttir *et al.*, 2008). The genetic variance of this trait was equal when assessed by the judges and the riders but the residual variance was much lower when assessed by the judges, giving a higher estimated heritability of the trait when assessed by the judges and indicating more reliability of these assessments. This can be attributed to the judges being more experienced in assessing the trait and are required to have a good knowledge of the judging scale and the definitions pertaining to each score, whereas the riders are more likely to base their assessments only on previous experience with received scores. Furthermore, few outliers in the dataset from the riders (score 5.0 for spirit) indicated a clouded judgement due to a personal frustration towards the horse presented. The genetic correlation between assessments made by the riders and the judges for spirit was also relatively low, supporting that there were differences between the groups in the use of the judging scale. The phenotypic correlation between assessments made by the riders and the judges for spirit is nevertheless higher than for the suppleness traits which may be explained with the novelty of the suppleness trait assessments, as spirit is more rooted within the assessment procedures and a more common understanding of the appearance of the trait have been established.

The estimated genetic variance for the suppleness traits differed depending on whether riders or judges assessed the trait. It seems that the riders were more able to assess traits such as cooperation, nerve strength and suppleness. This is logical, considering that riders are in direct physical contact with the horses while they are being assessed and therefore are in a better position than the observing judges to assess these traits. Moreover, in many cases the more experienced riders make efforts to hide any discrepancies in the performance of the horse pertaining to these traits and often succeed in doing so, which makes it more difficult for the judges to notice faults. These three traits are all related and are therefore difficult to distinguish, even for trained judges. For instance, a horse can be very supple and cooperative but lacks nerve strength which can mask the cooperativeness and suppleness under certain circumstances where the horse is insecure. However, estimated heritabilities tended to be higher for nerve strength and overall suppleness in this study compared with in the study by Birgisdóttir (2015). It may suggest that the judges are gaining experience in assessing these traits and to distinguish between them, and that these traits are not completely masked by the riders.

The range of heritabilities estimated for rein contact assessed by judges and riders were similar. It is logical to assume that the riders are better equipped to assess this trait, considering that they literally have the trait in their hands. The judges seem to apprehend similar genetic variance with their assessments for this trait, indicating validity of the assessments to a certain extent. However, the genetic and phenotypic correlations between the assessments of rein contact made by the riders and judges were rather low which may indicate differences in definition of the trait between the two groups. What the riders consider as good rein contact is not necessarily perceived as good rein contact by the judges and vice versa. Another aspect of this is the bits and bridles used. The riders choose a bit that is best suited for each horse, but it may be assumed that in some cases the chosen bit is not necessarily the best for the horse and the main goal of its use is to diminish apparent flaws pertaining to the rein contact. Therefore the rein contact can appear supple and light to the observing judge but in fact the horse is avoiding rein contact due to soreness or discomfort. The negative genetic correlation between rein contact and nerve strength when assessed by the judges partly supports this suggestion, as it can be assumed that the horse expresses some tension when experiencing discomfort. However, the genetic correlation between these two traits is ambiguous as it is positive when assessed by the judges in the sample group. Assessments for head carriage could also be influenced by the use of different bits, as it can be easier to bring the horse on the bit using these auxiliaries and thereby the head carriage may be forced to a certain extent but appearing good and stable. Certain types of bits have been reported to have considerable influence on bit-related lesions which was found to be a general problem in Icelandic competition horses (Björndóttir *et al.*, 2014). This has also been found relevant for breeding horses and subsequently, bits found to be a decisive risk factor have been banned in competitions and breeding field tests according to Icelandic Animal Welfare laws (no. 55/2013). However, it cannot be assumed that certain types of bits are the only reason for these issues as the force applied to the rein contact by the rider is highly relevant. The nature of rein contact gives an opportunity for objective assessment with the use of a dynamometer for example, but the execution of observations like those could be problematic in regards to cost and implementation of the method to the current procedures at breeding field tests. It may also be difficult to gain acceptance of these measures from the riders. The use of dynamometer could nevertheless be beneficial for validation of the assessment method for rein contact.

The results for head carriage and top line were rather unclear, and the fact that these traits were not assessed in 2015 made the dataset for analyses even smaller. The genetic variances varied greatly between analyses and were estimated with high standard errors in some cases which indicate that the results cannot be fully trusted. However, the results indicate that the traits comprise considerable genetic variation, especially when assessments made by the judges were analysed together in a bivariate analysis. This may indicate that these two traits at least partly describe the same trait. The question can be raised about whether an unbalanced horse displaying a stiff top line can in fact express a stable head carriage as, by nature, the horse uses the head and neck in order to gain balance. This is also supported by the high genetic correlation between these traits in the analysis for all assessed horses and relatively high phenotypic correlation when assessed by the riders. Therefore it may have been a better solution to combine these two traits instead of dismissing them in the suppleness assessment procedures in 2015. Moreover, the phenotypic correlations between assessments made by riders and judges for these two traits are the highest of all the suppleness traits, which indicates that the judges may be better equipped to assess these two traits than the other traits.

The riders and the judges perceive the temperament of the horse in different ways and under different circumstances. The PCA results indicate that the riders emphasise traits pertaining to physical constitution such as head carriage, rein contact and top line when assessing overall suppleness, but the judges emphasise traits pertaining to the horse's state of mind such as cooperation and nerve strength when assessing overall suppleness. The result is nevertheless ambiguous as overall suppleness loads on both principal components for both groups, which indicates that all the traits are taken into account by both groups when assessing overall suppleness. However, when assessing spirit both groups seem to emphasise traits pertaining to the horse's state of mind.

There is no indication that the assessments for the temperament traits are carried out in an insufficient way by the breeding judges. On the contrary the assessments are fairly normally distributed and there was genetic variance in some of the traits. However, with the current assessment method they have difficulties apprehending the genetic variance that seem to be present within the suppleness traits pertaining to horse's state of mind. It can be assumed that several factors contribute to this, where the most important ones are the subjective assessment method and perhaps lack of better definition of the traits, and the fact that rider may try to hide any discrepancies in the performance to improve the economic value of the horse. The addition of assessing the suppleness traits for the last two years has most likely resulted in better assessments of spirit by the judges, even though it is not evident in the results of this study. With the additional assessments the judges assess the spirit of the horse with different perspective where the overall suppleness of the horses is emphasised. It may therefore be interesting to continue the assessments for the suppleness traits, especially since the heritability estimates for the traits seem to have increased between the first two years. The current assessment procedure for spirit has resulted in a high estimated heritability of the trait and considerable genetic gain, which indicates that the trait is well defined within the breeding objective. Therefore adjustments like these to the current assessment method may be more beneficial than drastic changes in the breeding system. The information provided by the riders pertaining to the horse's state of mind is valuable considering the high heritability estimated for these traits. It is therefore of great importance to the breeding system that the riders contribute to better assessments by the judges with transparency in the presentation of the horses. The roles of horse owners and breeders are also of great significance because they are responsible for performing the selection of horses for breeding.

6.2 General temperament traits assessed in the everyday environment

Considering the importance of the trait spirit and the previous discussions about the reliability of the assessment method within the horse society, the response rate to the questionnaire survey was surprisingly low. The dataset only included about one fifth of all the horses that were assessed at a breeding field test in Iceland in 2014 and 2015. The inner structure of the questionnaire was described by four different principal components which pertained to traits of general nerve strength, cooperation, independence or sociability and training response, whereof traits included in general nerve strength and training response produced the most interesting and logical results. Most traits within general cooperation and independence/sociability were estimated with low heritability and high standard error values and were considered unreliable results. This may be due to different requirements and definitions of cooperation within the group of various assessors and perhaps reflect the different training levels of the horses assessed.

Traits pertaining to training response were time in training and training level, which were also investigated as fixed effects in the model used to estimate genetic parameters for the general temperament traits. Training level was found to have significant effect on many of the traits in the questionnaire and was included when estimating genetic parameters in bivariate analyses containing spirit. By including training level, genetic variance decreased a bit for some traits but increased for reaction to a novel/loud sound and spirit, resulting in a heritability for spirit close to what has been reported before (Árnason & Sigurðsson, 2004; Albertsdóttir *et al.*, 2008). The fixed effect of training level was however not included in the models used for genetic estimates in these studies. The greatest decrease of estimated heritability was within the trait behaviour while being trained (tension), where genetic variance decreased by half. It would be logical to assume that training would influence the tension factor in the horse, as many training methods involve desensitizing the horse and the rider gaining trustworthy leadership role in the horse-rider relationship. There is also the possibility that the same genes partly influence both the tension factor and response to training, and thereby some of the genetic variance of the trait would be removed by adjusting for training level. By those means it may be questioned if using training level as fixed effect is ideal.

Time in training and training level were also treated as biological traits to describe the genetic ability of the horse to respond to training. It is interesting to see that training level was estimated with a high heritability, and the estimates for both training level and time in training were higher than what has been estimated for trainability in other studies (von Borstel, 2013). This may indicate a genetic factor influencing how easily the horses can be trained, how fast they develop as a response to training and how soon they can be ready to be assessed at breeding field tests. This knowledge could be of importance for the breeding system. The aim for breeding easily trained horses that are ready to be tested at a young age could decrease the generation interval and increase genetic progress in the population. This can also be of economic importance as less work and time is needed before the horse can be presented at a breeding field test or competition and should therefore be considered in the breeding evaluation. However, the respondents to the questionnaire survey may have overestimated the training level of their horse, due to lack of knowledge of the training process, as indicated by the considerably high percentage of 4 year old horses said to have reached the highest training level. On the contrary, these few young horses may in fact be the ones that endow this genetic potential of developing fast as a response to training and were therefore ready to be presented at breeding field tests at such young age. Further research on the relationship between training level and time in training is needed to establish foundation for the integration of this trait in the breeding system. It would be interesting to investigate this in relation to the effects of age which is currently adjusted for in the breeding evaluation and in relation to repeated breeding assessments to establish a trend of training with time and improved performance.

Another interesting aspect of the results is the negative genetic correlation between traits in the questionnaire assessed in the everyday environment of the horse and spirit, especially the traits pertaining to general nerve strength and training response. The negative correlation between nerve strength traits assessed at home and score for spirit assessed at breeding field tests may explain the basis for the speculations about the breeding system favouring tense and nervous horses. Nevertheless, horses receiving high scores for nerve strength assessed at breeding field tests seem to have higher EBVs for spirit. This contradicts the findings of Brunberg *et al.* (2013) to a certain extent, but spirit and nerve strength are however moderately genetically correlated. There is also negative genetic correlation between spirit and time in training, which indicates that the horses that receive higher scores for spirit are among others the younger ones. This is not surprising though, as the youngest horses are often given the benefit of the doubt when being assessed for spirit by the breeding judges where the effort of the horse is emphasized more than the capability to perform. This may also indicate the necessity of good spirit for horses to be ready for being presented at breeding field tests at a young age. Taking into account the negative genetic relationship between spirit and both time in training and the nerve strength traits, one may wonder if energetic, apt horses that respond quickly to aids and training are considered nervous and tense horses by their owners.

Some of the traits included in the questionnaire survey were estimated with moderate to high heritabilities indicating a certain reliability of the assessments for these traits, which can be supported by the findings of Momozawa *et al.* (2003) that a questionnaire survey can be an effective way to assess temperament traits, especially those related to anxiety.

Low phenotypic correlations between the traits assessed in the questionnaire survey and the temperament traits assessed at breeding field tests is not surprising, as these assessments are carried out under entirely different circumstances. The current assessment procedures at breeding field tests for Icelandic horses are not adjusted for assessments like these included in the questionnaire survey, but as reported by Rothmann *et al.* (2014b), reactivity assessments from behavioural observations during conformation assessments and from owners and/or trainers based on a questionnaire concur and are correlated to performance traits. This indicates the possibility of adjusting the current procedures for observations like these during the conformation assessments for Icelandic horses to provide additional information to include in the assessment for spirit. Behaviour during the conformation assessment is currently used as a point of reference when assessing spirit as is stated in the judging scale. Nevertheless, more emphasis can be appointed to this part of the assessment and perhaps a simple linear scale could be of use to assess behavioural traits during the conformation assessment, which would only be used for reference in further assessment of spirit. Behavioural traits such as reactivity and handling for instance could be of interest, heritabilities of both traits have been estimated on a low to moderate range (0.17-0.28) for warmblood sport horses during a conformation assessment and thoroughbred racehorses during a routine veterinary examination (Rothmann *et al.*, 2014a; Oki *et al.*, 2007).

According to the results of the questionnaire survey, approximately 70% of the respondents concurred with the score their horse received for spirit when assessed at a breeding field test. The sample of horses assessed in the questionnaire survey had a slightly higher mean score for spirit than all the horse assessed at breeding field tests in 2014 and 2015. This may indicate that owners, breeders and/or trainers of horses that received high scores for spirit were more ready to respond to the survey which may give bias to these results. However, the low response rate to the survey may also indicate that a big part of horse owners and breeders consider the assessment of spirit acceptable and improvements of the assessment method not of importance. Therefore, it may be assumed that a big part of owners, breeders and trainers of Icelandic horses accept and concur with the scores given for spirit by breeding judges. The majority of the respondents who did not concur with the received score for spirit believed the score was underestimated. It cannot be stated though whether these respondents used the judging scale as a point of reference when expressing this argument or if they were merely stating their subjective opinion. It may therefore have been better to include the judging scale in the question to aim for more contemplated answers.

Suppleness was by far the most important attribute of the horse’s temperament according the respondents of the survey, and fiery was most frequently ranked as the least important attribute. Both attributes are used to describe the highest score for spirit, but fiery is not used to describe the lower scores and can therefore be thought of as a kind of a threshold for the highest score. A supple horse was defined as a cooperative horse that maintains light rein contact and seeks to please the rider, and a fiery horse was defined as an energetic and powerful horse, which are both characteristics of capable elite horses. However, it may be assumed that the respondents take caution with fiery, as the line between an energetic and powerful horse and a tense horse that is barely under control is ambiguous and may be one of the contributors to the speculations that tense and nervous horses are receiving high scores for spirit. This may therefore, suggest that a greater emphasis should be applied to the suppleness attribute in the description for the highest score for spirit in the judging scale. The current judging scale is old and the definitions of each score may therefore be irrelevant today as indicated by these results, but the Horse Breeders Association and FEIF have made plans to make adjustments to the judging scale in the forthcoming seasons.

The majority of the horses that were assessed in the survey were intended for breeding and/or competition in the future. It was not surprising that no respondent intended to cull the horse, as the horses assessed at breeding field tests are those that have passed a double preselection criteria, consisting of the preselection whether it should be subjected to continuing training after being broken in and the preselection whether it should be presented at a breeding field test.

6.3 Culling due to temperament faults

This part of the study demonstrated that estimated frequency of culling due to temperament faults is higher than has been estimated before in the Icelandic horse population (Björnsdóttir *et al.*, 2003; Sigfússon, 2003). Approximately one third of the horses in this study that were culled during the sample period were culled due to this reason. Over half of these horses were young horses that the owners decided to withdraw from training because of temperament problems. However, it cannot be assumed that temperament fault was the sole reason for culling as it was often combined with other reasons, such as conformation faults and lack of riding ability. In any case the majority of horses culled due to lack of riding ability were also young horses. Mean EBVs of young horses for spirit, riding ability and total score for conformation and riding ability did not differ between groups culled for different reasons. The sample included the period when the regular preselection is performed in which the decision is made whether a young horse is submitted to further training or not. Therefore, the high percentage of young horses in this sample may not be representative for the culling frequency of the entire year and may indicate some bias towards a younger age of the culled horses. The frequency of culling due to temperament faults could therefore be overrated.

Not many studies have focused on the causes of culling or death in horses and it can be assumed that the act of culling horses is not as common in other countries as it is in Iceland. The compatibility of culling rate in other breeds may therefore be questioned as it is uncommon to insure horses in Iceland, especially the less valuable ones, and there are no prerequisites for slaughtering horses. Nevertheless, culling rate of Swedish coldblood horses, mainly draught horses, due to temperament faults was reported to be relatively high (Wallin *et al.*, 2000). Both the Swedish coldblood horses and the Icelandic horses are known for their calm temperament which could be the result of rigorous selection methods consisting of the frequent culling due to temperament faults. A considerable percentage of culled horses during the sample period were culled due to high age, whereof the majority consisted of breeding horses and leisure horses. The high number of horses culled after the age of 20 years indicates that many horses stay healthy to a high age.

Further research within this area of culling horses would be of great interest, as the results indicate a high frequency of culling due to temperament faults of young horses. With a larger dataset covering a longer time period it may be possible to estimate genetic variance for this trait and perhaps include it in the breeding evaluation in the future. By this the breeding evaluation would be based on a higher proportion of the Icelandic horse population and the accuracy of the evaluation would increase. This trait could give additional information about the temperament of breeding horses, or their genetic potential to give well-tempered offspring, that cannot be apprehended at breeding field tests. The idea expressed at the symposium in 2013 to register the reason for culling in WorldFengur could therefore be of interest in order to collect data for further studies.

7 Conclusion

- The genetic variance of the new suppleness traits assessed for a trial period at breeding field tests are considerable and these traits provide additional information supporting the assessment of spirit. The assessment of these traits should therefore be continued and the re-introduction of the suppleness traits head carriage, top line and cooperation should be considered. Trait definitions need however to be improved.
- The score for spirit describes only part of the general temperament of the horse as the current assessment procedure is not adapted to assess traits pertaining to the horse's everyday environment.
- Response to training may be highly heritable trait that could contribute to breeding of apt horses that can be presented at breeding field tests at an early age, but further research is needed within the area.
- The majority of Icelandic horse owners, breeders and trainers in this study were satisfied with the assessment method for spirit.
- Approximately one third of culled horses were culled at least partly due to temperament faults whereof the majority were young horses that were withdrawn from further training, indicating a strong preselection based on temperament during the process of braking in young horses.
- The current breeding system does not apprehend the culling rate of horses with temperament faults and therefore further research is suggested to find ways to involve this information in the breeding system.

8 Acknowledgment

First of all, I would like to express my sincerest gratitude to my supervisors, Elsa Albertsdóttir and Susanne Eriksson. These extraordinary women have, with their endless knowledge, passion and patience, encouraged me and guided through the whole process of my work. Thank you for everything; it has been a privilege to have had the opportunity to work with you!

Without the help of the good people in the horse society in Iceland (and the horses of course!), this project would never have become reality. I would like to give special thanks to all the riders at the breeding field tests who contributed to this project as well as to the breeding judges for their tireless work. I would also like to thank Pétur Halldórsson for providing me with contact information about the riders. I gratefully thank all the horse owners, breeders and trainers that took the time to respond to the questionnaire survey and thereby providing valuable information about their horses. I would also like to give special thanks to all the owners of the culled horses for providing sensitive information and the staff of the slaughterhouses for their effort to help with the data collection.

I gratefully thank Þorvaldur Kristjánsson, the chairman of the Horse Breeders Association in Iceland, for his contribution to this project. He was most helpful with providing data from the breeding field tests and with the principal component analyses, as well as given me useful advices along the way.

I would like to thank all the horse experts who proofread the questionnaire for the survey and their valuable comments. These are Elsa Albertsdóttir, Einar Ásgeirsson, Freyja Þorvaldar, Hólmfríður Kristjánsdóttir, Mette Moe Mannseth, Tanja Rún Jóhannsdóttir and Þorvaldur Kristjánsson.

I would like to thank the Farmers Association in Iceland for hosting the questionnaire survey on the WorldFengur website. I would especially like to thank Þorberg Þorbergsson, the head of the IT department, for his pure kindness of taking the time to program the survey and make this possible.

I gratefully thank the Stock Protection Fund for the Icelandic horse and the Development Fund for Icelandic Horse Breeding for partly funding this project.

My thanks to the horse specific media in Iceland for their help in delivering messages to the people in the horse society when needed.

My sincere thanks to my examiner, Åsa Viklund, for her patience and interest in my work.

Last but certainly not least, I gratefully thank my family and friends for their support and patience during the process of this project. I would especially like to thank my parents for always having faith in me and supporting me in everything I do.

References

Aðalsteinsson, S. (1981). Origin and conservation of farm animal populations in Iceland. *Journal of Animal Breeding and Genetics, 98*(1-4), 258-264.

Agricultural University of Iceland. (2013, October 22). *Symposium on horse breeding (Málþing um kynbótakerfi í hrossarækt).* Retrieved from: http://www.lbhi.is/?q=is/malthing_um_kynbotakerfi_i_hrossaraekt_0

Albertsdóttir, E., Árnason, Th., Eriksson, S., Sigurðsson, Á., & Fikse, W. (2012). Effects of integrated genetic evaluations for Icelandic horses on predictive ability, accuracy and selection bias. *Journal of Animal Breeding and Genetics, 129*(1), 41-49.

Albertsdóttir, E., Eriksson, S., Näsholm, A., Strandberg, E., & Árnason, Th. (2007). Genetic analysis of competition data on Icelandic horses. *Livestock Science, 110*(3), 242-250.

Albertsdóttir, E., Eriksson, S., Näsholm, A., Strandberg, E., & Árnason, Th. (2008). Genetic correlations between competition traits and traits scored at breeding field-tests in Icelandic horses. *Livestock science, 114*(2), 181-187.

Albertsdóttir, E., Eriksson, S., Sigurðsson, Á., & Árnason, Th. (2011). Genetic analysis of 'breeding field test status' in Icelandic horses. *Journal of animal breeding and genetics, 128*(2), 124-132.

Andersson, L., & Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics, 5*(3), 202-212.

Árnason, Th. (1984). Genetic studies on conformation and performance of Icelandic toelterhorses IV. Best Linear Unbiased Prediction of ten correlated traits by use of an"Animal model". *Acta Agriculturae Scandinavica, 34*(4), 450-462.

Árnason, Th., & Sigurdsson, A. (2004). International genetic evaluations of the Icelandic horse. *55th Ann. Meet. EAAP, Bled, Slovenia,* 5-9.

Arnórsson, K. (2004). Ræktunin. In G. B. Björnsson, & H. J. Sveinsson, *Íslenski hesturinn* (pp. 202-247). Reykjavík: Mál og menning.

Arvelius, P., Asp, H., Fikse, W., Strandberg, E., & Nilsson, K. (2014). Genetic analysis of a temperament test as a tool to select against everyday life fearfulness in Rough Collie. *Journal of animal science, 92*(11), 4843-4855.

Axel-Nilsson, M., Peetz-Nielsen, P., Visser, E., Nyman, S., & Blokhuis, H. (2015). Perceived relevance of selected behavioural traits in horses – A survey conducted in Sweden. *Acta Agriculturae Scandinavica, 65*(1), 23-32.

Birgisdóttir, H. (2015). *Assessment of suppleness in Icelandic breeding horses (Mat á þjáni íslenskra kynbótahrossa).* (Bachelor thesis, Agricultural University of Iceland, Hvanneyri). Retrieved from: http://skemman.is/item/view/1946/22069

Björndóttir, S., Frey, R., Kristjánsson, Th., & Lundström, T. (2014). Bit-related lesions in Icelandic competition horses. *Acta Veterinaria Scandinavica, 56*(1), 1.

Björnsdóttir, S., Árnason, Th., & Lord, P. (2003). Culling rate of Icelandic horses due to bone spavin. *Acta Veterinaria Scandinavica, 44*(4), 1.

Björnsson, G. B., & Sveinsson, H. J. (2004). *Íslenski hesturinn.* Reykjavík: Mál og menning.

Brunberg, E., Gille, S., Mikko, S., Lindgren, G., & Keeling, L. (2013). Icelandic horses with silver coat colour show altered behaviour in a fear reaction test. *Applied Animal Behaviour Science, 146*(1), 72-78.

FEIF. (2016a). *FEIF Breeding Rules and Regulations*. Retrieved from: https://www.feif.org/files/documents/breeding2016_1.pdf

FEIF. (2016b). *Member Associations*. Retrieved from: https://www.feif.org/FEIF/MemberAssociations/Iceland.aspx

FEIF. (2016c). *Vision and Mission of FEIF*. Retrieved from: https://www.feif.org/FEIF/VisionandMission.aspx

Gille, C., & Spiller, A. (2010). Customer oriented horse breeding in Germany. Target group segmentation: an empirical analysis. *Züchtungskunde, 82*(3), 229-240.

Gille, C., Kayser, M., & Spiller, A. (2010). Target group segmentation in the horse buyers' market against the background of equestrian experience. *Journal of equine science, 21*(4), 67.

Górecka-Bruzda, A., Jastrzębska, E., Sosnowska, Z., Jaworski, Z., Jezierski, T., & Chruszczewski, M. (2011a). Reactivity to humans and fearfulness tests: Field validation in Polish Cold Blood Horses. *Applied Animal Behaviour Science, 133*(3), 207-215.

Górecka-Bruzda, A., Chruszczewski, M., Jaworski, Z., Golonka, M., Jezierski, T., Długosz, B., & Pieszka, M. (2011b). Looking for an ideal horse: Rider preferences. *Anthrozoös, 24*(4), 379-392.

Graf, P., von Borstel, U., & Gauly, M. (2013). Importance of personality traits in horses to breeders and riders. *Journal of Veterinary Behavior: Clinical Applications and Research, 8*(5), 316-325.

Hayek, A. R., Jones, B., Evans, D. L., Thomson, P. C., & McGreevy, P. D. (2005, August). Epidemiology of horses leaving the thoroughbred and standardbred racing industries. In *Proceedings of the 1st International Equitation Science Symposium, Broadford, Victoria, Post-Graduate Foundation in Veterinary Science, Sydney* (pp. 84-89).

Helgadóttir, Ú.Í. (2009). *Breeding assessments for Icelandic horses – Frequency of comments to riding ability scores and correlation between conformation and riding ability (Kynbótadómar íslenskra hrossa - Tíðni athugasemda hæfileika við einkunnir og tengsl byggingar og hæfileika).* (Bachelor thesis, Agricultural University of Iceland, Hvanneyri). Retrieved from: http://skemman.is/item/view/1946/7242

Horse Breeding Association council (2013). *Minutes from 65th council meeting December 17th 2013 (Fundargerð 65. fundar 17. desember 2013).* Retrieved from: http://www.rml.is/is/bufjarraekt/hrossaraekt/fundargerdir-fagrads

Hsu, Y., & Serpell, J. (2003). Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association, 223*(9), 1293-1300.

Icelandic Agriculture laws no. 70/1998.

Icelandic Animal Welfare laws no. 55/2013.

Indriðadóttir, E.Ó. (2012). *Second assessments at breeding field tests – frequency of score raising and estimation of genetic parameters (Yfirlitssýningar á kynbótahrossum – tíðni einkunnabreytinga og mat á erfðastuðlum).* (Bachelor thesis, Agricultural University of Iceland, Hvanneyri). Retrieved from: http://skemman.is/item/view/1946/13292

Koenen, E., Aldridge, L., & Philipsson, J. (2004). An overview of breeding objectives for warmblood sport horses. *Livestock Production Science, 88*(1), 77-84.

Kristjánsson, Th. (2001). *Evaluation of spirit in breeding horses (Mat á vilja og geðslagi kynbótahrossa).* (Unpublished bachelor thesis). Agricultural University of Iceland, Hvanneyri.

Kristjánsson, Th. (2014). *Riding ability in Icelandic horses – Effect of Conformation and the 'Gait keeper' mutation in the DMRT3 gene.* (Doctoral dissertation, Agricultural University of Iceland, Hvanneyri). Retrieved from: http://skemman.is/item/view/1946/20620

Madsen, P. & Jensen, J. (2013). *A User's Guide to DMU. Version 6, release 5.2.* University of Aarhus, Denmark. Retrieved from: http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf

Momozawa, Y., Kusunose, R., Kikusui, T., Takeuchi, Y., & Mori, Y. (2005a). Assessment of equine temperament questionnaire by comparing factor structure between two separate surveys. *Applied Animal Behaviour Science, 92*(1), 77-84.

Momozawa, Y., Ono, T., Sato, F., Kikusui, T., Takeuchi, Y., Mori, Y., & Kusunose, R. (2003). Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. *Applied Animal Behaviour Science, 84*(2), 127-138.

Momozawa, Y., Takeuchi, Y., Kusunose, R., Kikusui, T., & Mori, Y. (2005b). Association between equine temperament and polymorphisms in dopamine D4 receptor gene. *Mammalian Genome, 16*(7), 538-544.

Oki, H., Kusunose, R., Nakaoka, H., Nishiura, A., Miyake, T., & Sasaki, Y. (2007). Estimation of heritability and genetic correlation for behavioural responses by Gibbs sampling in the Thoroughbred racehorse. *Journal of Animal Breeding and Genetics, 124*(4), 185-191.

Pasing, S., & von Borstel, U.K. (2012, July). Evidence of grade inflation in personality trait scores from stallion performance tests. In *8th International Equitation Science Conference.*

Rothmann, J., Christensen, O., Söndergaard, E., & Ladewig, J. (2014a). A note on the heritability of reactivity assessed at field tests for Danish warmblood horses. *Journal of Equine Veterinary Science, 34*(2), 341-343.

Rothmann, J., Christensen, O., Söndergaard, E., & Ladewig, J. (2014b). Behavior Observation During Conformation Evaluation at a Field Test for Danish Warmblood Horses and Associations with Rideability and Performance Traits. *Journal of Equine Veterinary Science, 34*(2), 288-293.

SAS (Statistical Analysis Systems). (2016). *SAS 9.4 Product Documentation.* SAS Institute Inc., Cary, NC, USA. Retrieved from: http://support.sas.com/documentation/94/

Seaman, S., Davidson, H., & Waran, N. (2002). How reliable is temperament assessment in the domestic horse (Equus caballus)? *Applied Animal Behaviour Science, 78*(2), 175-191.

Sigfússon, V. (2003). *Longevity in Icelandic horses – length of life and reasons for culling and death.* (Unpublished master thesis). Swedish University of Agricultural Sciences, Uppsala.

Sigurðardóttir, H. (2012). *Investigation of selection intensity and genetic gain within breeding field test traits in Icelandic horse breeding (Athugun á úrvalsstyrkleika og erfðaframförum í einstökum eiginleikum í íslenskri hrossarækt).* (Bachelor thesis, Agricultural University of Iceland, Hvanneyri). Retrieved from: http://skemman.is/item/view/1946/13293

Wallin, L., Strandberg, E., Philipsson, J., & Dalin, G. (2000). Estimates of longevity and causes of culling and death in Swedish warmblood and coldblood horses. *Livestock production science, 63*(3), 275-289.

Visser, E., Van Reenen, C., Rundgren, M., Zetterqvist, M., Morgan, K., & Blokhuis, H. (2003). Responses of horses in behavioural tests correlate with temperament assessed by riders. *Equine veterinary journal, 35*(2), 176-183.

Visser, E., van Reenen, C., van der Werf, J., Schilder, M., Knaap, J., Barneveld, A., & Blokhuis, H. (2002). Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiology & Behaviour, 76*(2), 289-296.

von Borstel, U. (2013). Assessing and influencing personality for improvement of animal welfare: A review of equine studies. *CAB reviews, 8*(6), 1-27.

von Borstel, U., Euent, S., Graf, P., König, S., & Gauly, M. (2011a). Equine behaviour and heart rate in temperament tests with or without rider or handler. *Physiology & Behaviour, 104*(3), 454-463.

von Borstel, U., Pasing, S., & Gauly, M. (2011b). Towards a more objective assessment of equine personality using behavioural physiological observations from performance test training. *Applied Animal Behaviour Science, 135*(4), 277-285.

von Borstel, U., Pirsich, W., Gauly, M., & Bruns, E. (2012). Repeatability and reliability of scores from ridden temperament tests conducted during performance tests. *Applied Animal Behaviour Science, 139*(3), 251-263.

WorldFengur - The Studbook of Origin for the Icelandic horse. (2016). Retrieved from: http://www.worldfengur.com

Appendix I

Judging scale for spirit:

9.5 -10.0:
The horse should be fiery, cheerful and brave, but extremely easy to handle. All the time, the horse tries to please the rider

9.0:
- Very willing and eager but sensible and easy to handle, not fiery
- Very willing, but only fairly sensible and easy to manage
- Very eager to please and cooperate, but is not fiery

8.5:
- Very sensible, but not extremely forward going
- Very willing and eager, but only fairly sensible and easy to handle

8.0:
- Pleasantly willing when ridden
- Very willing and eager, but not easy to handle

7.5:
- Sensible and easy to handle, but not forward going
- Willing but stressed
- Pleasantly willing when ridden but sensitive or unfocused

7.0:
- Not willing or forward going
- Shows disobedience
- Nervous

6.5-5.0:
- Disobedient
- Lazy and dull
- Uncontrollable (bolting)

Appendix II

Judging scale for suppleness:

7.0:
The horse is extremely supple, light and cooperative in every aspect. It is apt, alert, focused, courageous and completely free of tension. Rein contact is light and supple in all gaits and the horse is able and willing to carry it self correctly with a supple top line and free of stiffness.

6.0:
- The horse expresses great suppleness. Rein contact is light and the top line is supple and free of stiffness.
- The horse is very supple and cooperative, free of tension and responds to the aids with considerable ease.
- The horse responds to all aids with great ease, is apt and free of tension but a minor instability characterises its performance.

5.0:
- The horse is rather supple and free of tension, head carriage is quite stable but rein contact and response to aids could be lighter and met with more ease.
- The horse is supple and cooperative but tension and/or insubordination can be detected in few occasions.

4.0:
- The horse expresses average suppleness.
- The horse is rather supple but tension and/or stiffness can be detected in the horse's top line.
- The horse responds to aids with a minor insubordination but is free of tension.
- The horse responds to aids with ease but expresses some nervousness and fearfulness.

3.0:
- The horse is uncooperative.
- The horse responds to aids with some insubordination.
- The horse is stiff or heavy on the reins.
- The horse expresses nervousness and fearfulness.
- Considerable tension and/or stiffness is detectable in the horse's top line.

2.0:
- The horse is substantially uncooperative.
- The horse responds to aids with substantial insubordination.
- The horse is never light on the reins and expresses great stiffness.
- The top line (and/or the mouth) of the horse is very stiff and tense.
- The horse expresses a sign of stubbornness or being out of control.
- The horse expresses obvious nervousness and/or fearfulness.

1.0:
- The horse is out of control, stubborn and attempts to bolt or rear/buck.
- The horse is not able to finish the test due to commotion or decisive cooperation faults

If the horse expresses obvious temperament faults during the conformation assessment (for example lack of cooperation, tension, coldness, nervousness) it is used for suppleness assessment in the ridden ability assessment.

Appendix III

Spirit – Data collection at breeding field tests in 2014

Breeding field test at Gaddstaðaflatir, June 2nd – 6th and 10th – 13th 2014
Horse / Origin / ID-number: ____
Date:______________________________

1. Has the rider also been training the horse being tested?
   (   ) Yes   (   ) No

2. If so, for how long has the horse been trained for the test?
   (   ) Less than a month
   (   ) 1-2 months
   (   ) 3-5 months
   (   ) 6 months or more

3. Is the rider also the owner of the horse being tested?
   (   ) Yes   (   ) No

4. Which score for spirit would the horse receive according to the rider based on performance during the test (scale 5-10)
   ____

5. Assessment for suppleness

Rider assesses the horse on a linear scale 1-7 for 5 traits pertaining to suppleness (mark the appropriate score in table). Rider also assesses a total score for suppleness on a linear scale 1-7, based on the judging scale below.

Five traits describing suppleness:

Trait		1	2	3	4	5	6	7	
Head carriage	Unstable head carriage								Stable head carriage
Rein contact	Stiff/heavy								Light/supple
Top line	Stiff								Supple
Cooperation	Uncooperative Aggressiveness Out of control								Cooperative Lightness
Nerve strength	Tenseness Nervousness Fearfulness								Composed Determination Courage

Overall suppleness (1-7): ______________

Spirit – Data collection at breeding field tests in 2015

Breeding field test at Gaddstaðaflatir, June 8th – 12th and July 20th – 25th 2015
Horse / Origin / ID-number: ____
Date:_________________________________

1. Has the rider also been training the horse being tested?
   ( ) Yes ( ) No

2. If so, for how long has the horse been trained for the test?
   ( ) Less than a month
   ( ) 1-2 months
   ( ) 3-5 months
   ( ) 6 months or more

3. Is the rider also the owner of the horse being tested?
   ( ) Yes ( ) No

4. Which score for spirit and trait description would the horse receive according to the rider based on performance during the test (scale 5-10)

   ____

- ☐ Fiery, indefatigable
- ☐ Eager, enthusiastic
- ☐ Willing
- ☐ Cooperative
- ☐ Alert
- ☐ Lazy
- ☐ Out of control
- ☐ Tenseness
- ☐ Aggressiveness
- ☐ Cold
- ☐ Stubborn
- ☐ Nervous
- ☐ Apathetic, dull
- ☐ Uncooperative
- ☐ No enterprise

5. Assessment for suppleness

Rider assesses the horse on a linear scale 1-7 for 2 traits pertaining to suppleness, i.e. rein contact and nerve strength (mark the appropriate score in table). Rider also assesses a total score for suppleness on a linear scale 1-7, based on the judging scale below.

Two traits describing suppleness:

Trait		1	2	3	4	5	6	7	
Rein contact	Stiff/heavy								Light/supple
Nerve strength	Tenseness								Composed

Overall suppleness (1-7): ________________

Appendix IV

Assessment of temperament traits made by owner/breeder/trainer for breeding horses in their daily environment

Name and origin of the horse: ______________________________

ID-number.: ______________________________

Test year: ☐ 2014 ☐ 2015

How is the horse connected to the respondent?

Multiple options are allowed.

- ☐ Breeder
- ☐ Owner
- ☐ Co-owner
- ☐ Trainer
- ☐ BFT rider
- ☐ Other: ______________________________

<u>NB, questions below should only be answered for horses previously assessed in a breeding field test in 2014 or 2015</u>

<u>Unless otherwise specified, only one choice is allowed in the questions below</u>

1. For how long has the horse been in training?

Refers to from the beginning when the horse was broken in until it was assessed in a breeding field test in 2014 or 2015.

1 ☐ Less than 12 months

2 ☐ 12-24 months

3 ☐ 25-36 months

4 ☐ More than 36 months

☐ Unknown

2. Which training level has the horse reached (acc. to the Icelandic training ladder by Holar University)?

1 ☐ Level 1

The horse has been subjected to a basic training and reached the first three levels of the Icelandic training ladder where the emphasis is on the importance of forward thinking, mental stability, free and uninhibited movements and good balance in all gaits.

2 ☐ Level 2

The horse has received considerable training and reached the fourth and fifth level of the Icelandic training ladder where the emphasis is on good balance in all gaits, symmetric and straight movements with steady rein contact and long, supple top line free of tension.

3 ☐ Level 3

The horse has received extensive training and reached the last three training levels of the Icelandic training ladder where the emphasis is on energetic and powerful horse that is able to perform in a collected posture with good self-carriage and strength to perform in fast speed.

☐ Unknown

3. How does the horse react when seeing novel objects?
Refers to circumstances where the horse sees a novel object for the first time in its daily environment.

1 ☐ Very nervous
2 ☐ Nervous
3 ☐ Calm
4 ☐ Very calm
☐ Unknown

4. What are the horse's reactions when seeing novel objects?
Refers to circumstances where the horse sees a novel object for the first time in its daily environment without support from humans. Please mark the option that best describes the first reaction of the horse.

1 ☐ Gets scared and flees instantly
2 ☐ Lingers and then flees
3 ☐ Lingers and then approaches
4 ☐ Approaches instantly
☐ Unknown

5. How does the horse react when hearing novel and/or loud sound?
Refers to circumstances where the horse hears a novel/loud sound in its daily environment.

1 ☐ Very nervous
2 ☐ Nervous
3 ☐ Calm
4 ☐ Very calm
☐ Unknown

6. How does the horse react when removed temporarily from other horses?
For example when the horse is ridden without the support of other horses or any kind of handling where other horses are not visible.

1 ☐ Very nervous
2 ☐ Nervous
3 ☐ Calm
4 ☐ Very calm
☐ Unknown

7. How does the horse react to new environment with other horses present?
For example when the horse is moved to a new stable or is ridden in a new environment.

1 ☐ Very nervous
2 ☐ Nervous
3 ☐ Calm
4 ☐ Very calm
☐ Unknown

8. How does the horse behave towards humans in its daily environment?
Refers to behaviour in general association with humans, not during training.

1 ☐ Very nervous
2 ☐ Nervous
3 ☐ Calm
4 ☐ Very calm
☐ Unknown

9. How does the horse behave towards other horses in its daily environment?
Refers mainly to reaction towards unfamiliar horses, for example when a new horse arrives at the stable or is introduced to a group of horses.

1 ☐ Very aggressive
2 ☐ Aggressive
3 ☐ Friendly
4 ☐ Very friendly
☐ Unknown

10. How does the horse react to disturbance while being trained?
Disturbance can mean a sudden environmental sound, sudden movement (e.g. bird flying up), traffic by car or other riders etc.

1 ☐ Very unassured
2 ☐ Unassured
3 ☐ Assured
4 ☐ Very assured
☐ Unknown

11. How did the horse react to new aids when he was broken in and at the beginning of training?
Refers to aids used for general handling, for riding and for training by hand. Please mark appropriate option for unapt/apt on the one hand and for tenseness on the other hand.

1 ☐ Very unapt
2 ☐ Unapt
3 ☐ Apt
4 ☐ Very apt
☐ Unknown

1 ☐ Very tense
2 ☐ Tense
3 ☐ Relaxed
4 ☐ Very relaxed
☐ Unknown

12. How does the horse normally behave while being trained?
Refers to behaviour when ridden and trained by hand. Please mark appropriate option for cooperation on the one hand and for tenseness on the other hand.

1 ☐ Very uncooperative
2 ☐ Uncooperative
3 ☐ Cooperative
4 ☐ Very cooperative
☐ Unknown

1 ☐ Very tense
2 ☐ Tense
3 ☐ Relaxed
4 ☐ Very relaxed
☐ Unknown

13. How ready and willing is the horse when ridden away from the stable?
Refers to circumstances when the horse is ridden away from the stable without the support of another horse.

1 ☐ Very unwilling
2 ☐ Unwilling
3 ☐ Willing
4 ☐ Very willing
☐ Unknown

14. How cooperative is the horse in general?
Refers to cooperation in general handling, when ridden and when trained by hand.

1 ☐ Very uncooperative
2 ☐ Uncooperative
3 ☐ Cooperative
4 ☐ Very cooperative
☐ Unknown

15. How predictable is the horse in general?
Predictable horse responds to similar circumstances in a similar way. Unpredictable horse responds to similar circumstances in different ways.

1 ☐ Very unpredictable
2 ☐ Unpredictable
3 ☐ Predictable
4 ☐ Very predictable
☐ Unknown

16. How well does the score for *Spirit* conform to the general temperament of the horse?

1 ☐ Very badly
2 ☐ Badly
3 ☐ Well
4 ☐ Very well
☐ Unknown

If badly or very badly, how is the inconsistency explained?

☐ Overestimated
☐ Underestimated
Explanation: ________________________________

17. What temperament trait is the most important according to the respondent?
Please rank the traits according to importance by numbering 1 as the most important one, 2 the second most important one etc. Ranking all traits is not obligatory.

☐ Suppleness – *the horse is cooperative, light on the reins and seeks to please the rider*
☐ Eager – *the horse has good forward thinking under complete control of the rider*
☐ Nerve strength – *the horse is free of tension/stress and is courageous*
☐ Apt – *the horse is quick to learn and responds well to light aids*
☐ Hard working – *the horse is selfless and committed to any task*
☐ Fiery – *the horse is happy, energetic and powerful*
☐ Other: ________________________________

18. What is the intended future role of the horse?
Multiple options are allowed.

- ☐ Breeding horse
- ☐ Riding/leisure horse
- ☐ Competition horse
- ☐ For sale
- ☐ Culling
- ☐ Other: ___________________________________

Thank you for your participation!

Appendix V

Table 1. Analysis of variance of different general linear models for Part I considering significance ($p<0.05$) of different fixed effects (age_sex, field test, test year and rider group) and coefficient of determination (R^2); - indicates non-significant effect

	Y = age_sex + field test			**Y = age_sex + test year**			**Y = age_sex + field test + rider group**				**Y = age_sex + test year + rider group**			
	age_sex	field test	R^2	age_sex	test year	R^2	age_sex	field test	rider group	R^2	age_sex	test year	rider group	R^2
Assessed by judges														
Spirit	-	-	4%	-	-	3%	-	-	-	5%	-	-	0.02	5%
Head carriage	0.03	-	7%	0.03	*	7%	0.03	-	0.04	10%	0.03	*	0.03	10%
Rein contact	-	<.0001	9%	-	<.0001	6%	-	<.0001	0.01	11%	-	<.0001	0.04	7%
Top line	-	-	5%	-	*	5%	-	-	-	6%	-	*	-	6%
Cooperation	0.03	-	7%	0.04	*	6%	0.03	-	-	7%	0.04	*	-	7%
Nerve strength	-	-	4%	-	-	3%	-	-	-	5%	-	-	-	4%
Suppleness total	0.02	0.04	6%	0.02	-	4%	0.03	0.02	-	7%	0.02	-	-	5%
Assessed by riders														
Spirit	0.04	-	4%	0.03	-	4%	0.03	-	-	4%	0.03	-	-	4%
Head carriage	0.02	-	8%	0.02	*	7%	0.03	-	-	8%	0.02	*	-	7%
Rein contact	0.00	0.01	8%	0.00	-	5%	0.00	0.01	-	8%	0.002	-	-	5%
Top line	-	0.05	4%	-	*	3%	-	-	-	4%	-	*	-	3%
Cooperation	-	-	5%	-	*	5%	-	-	-	7%	-	*	-	6%
Nerve strength	<.0001	-	8%	<.0001	0.04	8%	<.0001	-	-	8%	<.0001	0.04	-	8%
Suppleness total	0.01	0.004	8%	0.01	0.02	6%	0.01	0.01	-	8%	0.01	0.02	-	7%

* Head carriage, top line and cooperation were not assessed in 2015 and therefore the effect of test year does not apply for these traits

Table 2. Analysis of variance of different general linear models for Part II considering significance ($p<0.05$) of different fixed effects (age_sex, training time and training level) and coefficient of determination (R^2); - indicates non-significant effect

	Y = age_sex		Y = age_sex + training time			Y = age_sex + training level		
	age_sex	R^2	age_sex	Training time	R^2	age_sex	Training level	R^2
Q1 Time in training	<.0001	50%				<.0001	<.0001	55%
Q2 Training level	<.0001	18%	0.03	<.0001	28%			
Q3 Behaviour when seeing a novel object	0.01	5%	0.01	-	7%	0.03	-	6%
Q4 Reaction to a novel object	-	4%	0.04	-	6%	0.07	-	5%
Q5 Reaction to novel/loud sound	0.04	4%	0.04	-	6%	0.06	0.04	7%
Q6 Reaction to temporary isolation	-	3%	-	-	5%	-	0.03	5%
Q7 Reaction to new environment	-	2%	-	-	3%	-	-	3%
Q8 Behaviour towards humans	0.02	5%	0.01	-	7%	0.01	-	7%
Q9 Behaviour towards other horses	-	3%	-	-	5%	-	-	5%
Q10 Reaction to disturbance while being trained	0.00	7%	0.00	-	8%	0.00	0.04	9%
Q11a Reaction to new aids while being broken in (apt/unapt)	-	3%	-	-	4%	-	-	4%
Q11b Reaction to new aids while being broken in (tension)	0.02	5%	0.01	-	6%	0.03	-	7%
Q12a Behaviour while being trained (cooperation)	-	2%	-	-	2%	-	0.03	4%
Q12b Behaviour while being trained (tension)	0.01	6%	0.00	-	8%	0.02	0.00	10%
Q13 Behaviour while being ridden away from stable	-	3%	-	-	6%	-	0.01	6%
Q14 General cooperation in handling	-	1%	-	-	3%	-	0.00	6%
Q15 Behaviour predictability	-	3%	-	-	3%	-	-	5%

Appendix VI

Table 1. Heritabilities (h^2), additive genetic (σ^2_a) and residual (σ^2_e) variances (standard errors as subscripts) estimated in bivariate analyses with the trait spirit (using model 3) for traits assessed by owners/breeders/trainers in the survey, and genetic correlation (r_g) (standard errors as subscripts) and phenotypic correlation (r_p) between spirit and the same traits; heritabilities in bold indicate significant results

	h^2	σ^2_a	σ^2_e	r_g	r_p
Spirit	$0.24_{0.177}$	$0.04_{0.030}$	$0.13_{0.029}$		
General nerve strength component					
Q3 Reaction to a novel object	$0.10_{0.139}$	$0.05_{0.077}$	$0.50_{0.082}$	$0.59_{0.814}$	-0.03
Q4 Action when seeing a novel object	$0.29_{0.191}$	$0.14_{0.096}$	$0.34_{0.089}$	$-0.02_{0.524}$	-0.04
Q5 Reaction to novel/loud sound	$0.22_{0.159}$	$0.11_{0.084}$	$0.40_{0.082}$	$-0.17_{0.522}$	0.01
Q7 Reaction to new environment*	$0.09_{0.157}$	$0.04_{0.069}$	$0.40_{0.072}$	$0.87_{0.932}$	0.04
Q8 Behaviour towards humans	$0.11_{0.141}$	$0.05_{0.058}$	$0.36_{0.061}$	$0.11_{0.733}$	0.03
Q10 Reaction to disturbance while being trained	$0.21_{0.176}$	$0.12_{0.100}$	$0.44_{0.096}$	$-0.01_{0.577}$	-0.02
Q11b Reaction to new aids while being broken in (tension)	$0.29_{0.191}$	$0.14_{0.100}$	$0.36_{0.094}$	$-0.53_{0.455}$	-0.01
Q12b Behaviour while being trained (tension)	$0.20_{0.169}$	$0.08_{0.068}$	$0.32_{0.066}$	$0.16_{0.576}$	0.05
Q15 Behaviour predictability	$0.00_{0.151}$	$0.00_{0.049}$	$0.32_{0.053}$	$1.00_{26.60}$	0.11
General cooperation component					
Q11a Reaction to new aids while being broken in (apt/unapt)	$0.03_{0.145}$	$0.01_{0.062}$	$0.41_{0.068}$	$1.00_{2.225}$ [a]	0.05 [a]
Q12a Behaviour while being trained (cooperation)	$0.13_{0.157}$	$0.05_{0.061}$	$0.33_{0.062}$	$0.82_{0.692}$	0.19
Q13 Behaviour while being ridden away from stable*	$0.01_{0.133}$	$0.01_{0.050}$	$0.37_{0.056}$	$1.00_{4.758}$ [a]	0.12 [a]
Q14 General cooperation in handling	$0.03_{0.119}$	$0.01_{0.046}$	$0.38_{0.053}$	$1.00_{2.692}$ [a]	0.17
Independence/sociability component					
Q6 Reaction to temporary isolation	$0.02_{0.129}$	$0.01_{0.054}$	$0.41_{0.060}$	$1.00_{3.086}$ [a]	0.02 [a]
Q9 Behaviour towards other horses	$0.26_{0.177}$	$0.11_{0.079}$	$0.32_{0.074}$	$0.87_{0.396}$	0.08
Training response component					
Q1 Time in training	$0.21_{0.199}$	$0.09_{0.093}$	$0.36_{0.089}$	$-0.39_{0.631}$	0.12
Q2 Training level	$\mathbf{0.56_{0.189}}$	$0.23_{0.085}$	$0.18_{0.073}$	$0.44_{0.348}$	0.28

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)

* Reaction to new environment and behaviour while being ridden away from stable are also included in the independence/sociability component

Table 2. Genetic correlations (below the diagonal; standard errors as subscripts), average heritability (diagonal) and phenotypic correlations (above the diagonal) estimated in bivariate analyses using model 3 for traits loading on general nerve strength component assessed by owners/breeders/trainers in the survey; genetic correlations in bold indicate significant results

	Q3	Q4	Q5	Q7	Q8	Q10	Q11b	Q12b	Q15
Q3	0.12	0.55	0.64	0.53	0.48	0.64	0.50	0.49	0.43
Q4	$0.29_{0.601}$	0.29	0.59	0.37	0.42	0.61	0.38	0.37	0.35
Q5	$-0.08_{0.660}$	$0.60_{0.343}$	0.22	0.48	0.49	0.68	0.45	0.47	0.41
Q7	$-0.01_{0.857}$	$0.17_{0.619}$	$1.00_{0.579}$ [a]	0.12	0.43	0.50	0.36	0.45	0.41
Q8	$1.00_{0.565}$ [a]	$0.36_{0.546}$	$0.87_{0.499}$	$1.00_{0.673}$ [b]	0.13	0.46	0.44	0.45	0.35
Q10	$\mathbf{0.87_{0.317}}$	$\mathbf{0.85_{0.222}}$	$\mathbf{1.00_{0.283}}$ [b]	$0.52_{0.549}$	$\mathbf{1.00_{0.455}}$ [b]	0.21	0.51	0.52	0.45
Q11b	$0.07_{0.664}$	$0.36_{0.449}$	$0.65_{0.421}$	$0.68_{0.497}$	$0.43_{0.555}$	$\mathbf{0.72_{0.355}}$	0.28	0.66	0.36
Q12b	$0.53_{0.557}$	$0.41_{0.464}$	$\mathbf{1.00_{0.484}}$ [a]	$1.00_{0.522}$ [a]	$\mathbf{1.00_{0.438}}$ [b]	$\mathbf{1.00_{0.397}}$ [b]	$\mathbf{0.89_{0.228}}$	0.20	0.44
Q15	$-1.00_{86.11}$ [a]	$1.00_{1.24}$ [a]	$1.00_{30.05}$ [a]	$1.00_{2.449}$ [a]	$-1.00_{5.153}$ [a]	$1.00_{19.73}$ [a]	$-1.00_{2.731}$ [b]	$1.00_{9.093}$ [b]	0.02

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)
[b] Convergence criteria used was lower (norm vector of $<10^{-5}$ instead of $<10^{-7}$)

Q3 = Reaction to a novel object, Q4 = Action when seeing a novel object, Q5 = Reaction to novel/loud sound, Q7 = Reaction to new environment, Q8 = Behaviour towards humans, Q10 = Reaction to disturbance while being trained, Q11b = Reaction to new aids while being broken in (tension), Q12b = Behaviour while being trained (tension) and Q15 = Behaviour predictability

Table 3. Genetic correlations (below the diagonal; standard errors as subscripts), average heritability (diagonal) and phenotypic correlations (above the diagonal) estimated in bivariate analyses using model 3 for traits loading on the general cooperation component assessed by owners/breeders/trainers in the survey

	Q11a	Q12a	Q13	Q14
Q11a Reaction to new aids while being broken in (apt/unapt)	0.09	0.40	0.31	0.38
Q12a Behaviour while being trained (cooperation)	$0.14_{0.853}$	0.18	0.27	0.52
Q13 Behaviour while being ridden away from stable	$-1.00_{5.687}$ [a]	$-1.00_{90.68}$	0.03	0.46
Q14 General cooperation in handling	$-0.93_{1.362}$	$0.13_{0.957}$	$-1.00_{15.80}$ [a]	0.06

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)

Table 4. Genetic correlations (below the diagonal; standard errors as subscripts), average heritability (diagonal) and phenotypic correlations (above the diagonal) estimated in bivariate analyses using model 3 for traits loading on the independence/sociability component assessed by owners/breeders/trainers in the survey; genetic correlations in bold indicate significant results

	Q6	Q7	Q9	Q13
Q6 Reaction to temporary isolation	0.03	0.44	0.26	0.15
Q7 Reaction to new environment	$1.00_{14.92}$ [a]	0.12	0.31	0.28
Q9 Behaviour towards other horses	$1.00_{1.049}$ [a]	$\mathbf{1.00_{0.477}}$ [a]	0.25	0.20
Q13 Behaviour while being ridden away from stable	$1.00_{4.146}$ [a]	$1.00_{4.463}$ [a]	$1.00_{10.05}$ [a]	0.03

[a] Convergence criteria used was lower (norm vector of $<10^{-6}$ instead of $<10^{-7}$)