

ACCURACY OF HORSE AFFECT ASSESSMENTS

**ACCURACY OF HORSE AFFECT ASSESSMENTS: A COMPARISON OF
EQUINE ASSISTED MENTAL HEALTH PROFESSIONALS, NON-EQUINE
ASSISTED MENTAL HEALTH PROFESSIONALS, AND LAYPEOPLE**

SEBASTIAN A. FOX
Bachelor of Arts (Honours), Mount Royal University, 2020

A thesis submitted
in partial fulfillment of the requirements for the degree of

MASTER OF EDUCATION

in

COUNSELLING PSYCHOLOGY

Faculty of Education
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

© Sebastian A. Fox, 2023

ACCURACY OF HORSE AFFECT ASSESSMENTS

ACCURACY OF HORSE AFFECT ASSESSMENTS: A COMPARISON OF EQUINE ASSISTED MENTAL HEALTH PROFESSIONALS, NON-EQUINE ASSISTED MENTAL HEALTH PROFESSIONALS, AND LAYPEOPLE

SEBASTIAN A. FOX

Date of Defence: October 12th, 2023

Dr. Elaine Greidanus Thesis Supervisor	Associate Professor	Ph.D., R.Psych.
---	---------------------	-----------------

Dr. Louise Barrett Thesis Examination Committee Member	Professor of Psychology	Ph.D., F.R.S.C.
---	-------------------------	-----------------

Dr. Richelle Marynowski Thesis Examination Committee Member	Associate Dean, Professor	Ph.D.
--	---------------------------	-------

Dr. Holli-Anne Passmore External Examiner Concordia University of Edmonton Edmonton, Alberta	Associate Professor, Department Chair	Ph.D.
---	---------------------------------------	-------

Dr. Lorraine Beaudin Chair, Thesis Examination Committee	Associate Professor	Ph.D.
---	---------------------	-------

Dedication

This thesis is dedicated to my parents, whose support of my interest in horses as a young child brought my beloved Boy into my life. To my mother who sacrificed so many hours taking me to the stable, and to my father who was always encouraging.

And to my loving partner, who has been so supportive and whose brilliance helped tremendously in the completion of this thesis.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Abstract

In order to know whether a horse's welfare has been compromised, the professional needs to be able to assess the animal's behaviour and meaningfully interpret it. Misinterpretation of a horse's mental state may lead to further misunderstandings of the animal's behaviour, breakdowns in horse-human communication, and jeopardized wellbeing (Gronqvist et al., 2016; see also Horseman et al., 2016; McBride & Long, 2001; Merkies et al., 2018). This study investigated the abilities of (EAMHPs) to assess the affective states of horses.

First, an instrument for measuring accuracy was developed by showing a panel of 12 equine behaviour experts 20 videos of horses and asking them to describe the emotional states of the animals. Using the QBA and FCP method, the 10 videos with the highest inter-observer consensus and their associated terms (used to create answer keys) were retained. In the prediction testing phase, those 10 videos were shown to EAMHPs ($n = 55$), laypeople ($n = 94$), and non-equine assisted mental health professionals (NEAMHPs; $n = 51$), who were also asked to generate affectively descriptive terms. These three groups were then graded using the answer keys and awarded total assessment accuracy scores representing how similar their answers were to those of the experts. The participants of the three groups also self-rated their perceived level of horse experience and filled out a related questionnaire.

The results found that EAMHPs and laypeople scored significantly higher than the NEAMHPs. However, when horse experience scores were controlled for, the EAMHPs no longer scored significantly higher than the NEAMHPs. Profession and horse experience scores significantly accounted for variation in the total assessment accuracy

ACCURACY OF HORSE AFFECT ASSESSMENTS

scores, and the horse experience scores were positively correlated with self-rated levels of horse experience. Those participants who believed they had a high level of horse experience scored significantly higher than those who said they had no horse experience or a low level of experience.

The finding that EAMHPs and NEAMHPs score similarly, yet laypeople (significantly) outperform the latter and the former (non-significantly) when assessing the affective states of horses is perplexing. More research is needed to further investigate the involvement of horses in equine assisted therapy practices to ensure that their use remains ethical and that their welfare is not being compromised in exchange for aiding clients.

Keywords: equine-assisted therapy, equine, affect, emotional state, accuracy, assessment accuracy, horse, co-therapist, animal-assisted therapy, equine assisted mental health professionals, non-equine assisted mental health professionals, laypeople

Acknowledgements

I want to say thank you to my supervisor, Dr. Elaine Greidanus whose early support of this research made this thesis possible, and I wish to extend a thank you to my committee members Dr. Louise Barret and Dr. Richelle Marynowski, and my external examiner Dr. Holli-Anne Passmore.

I also want to extend my deepest gratitude to Nemanja Miltenovic who developed the codes used to prepare the data collected in this study for statistical analysis, and to my dear friend Catherine Lempke for aiding me in the proof-reading of this thesis.

Table of Contents

Literature Review4

A Historical Overview of Animal Assisted Therapy 4

 The Healing Properties of Animals.....5

 The Emergence of Animal Therapy.....5

 York Retreat.....6

 Corporal Wayne and Smoky.....6

 Freud and Jofi.....6

 Levinson and Jingles.....7

Horses in Psychology..... 8

 Therapeutic Traits of Horses.....8

 How Horses Help Humans.....9

 Physically.....9

 Cognitively.....9

 The Horse as Co-Therapist10

 Types of Equine Assisted Therapy.....11

 Associations.....11

 Summary.....12

Concerns of Equine Welfare 12

 Defining Affect12

 Negative Affect.....13

 Positive Affect.....13

ACCURACY OF HORSE AFFECT ASSESSMENTS

A Brief Overview of The Five Freedoms Paradigm.....	13
Physical Domain.....	15
Mental Domain.....	15
A Brief Overview of The Five Domains Model.....	15
Contributors to Equine Welfare Risk.....	16
Training and Riding.....	17
Training History.....	17
Lack of Knowledge.....	18
Consequences.....	18
Equine Welfare in EAT	19
Attitudes and Beliefs of Equine Assisted Professionals.....	19
Education is Important.....	20
Therapy Can be Unsafe For Horses.....	20
Horses Mostly Enjoy being Co-Therapists.....	20
EAT is Somewhat Adequately Regulated.....	21
Summary.....	21
Impacts of Client Populations on Equines.....	22
Youth.....	22
Experienced Riders.....	22
Clients with PTSD.....	23
Summary.....	23
<i>Making Meaning of Horse Behaviour</i>	<i>24</i>

ACCURACY OF HORSE AFFECT ASSESSMENTS

The Importance of Clear Communication	24
Horses Trying to Communicate with Humans.....	25
Implications for Equine Assisted Professionals.....	25
The Influence of Horse Experience	25
Lack of Experience Leads to Misinterpretations.....	26
How a Human’s Horse Experience Impacts the Animal.....	26
Summary.....	27
Significance to Equine Assisted Therapy	27
<i>Speaking Another Language: Can People Truly Understand Horses?.....</i>	28
What is Clear Communication?	28
How Emotion Impacts Animal Behaviour.....	28
Are People any Good at Assessing Animal Affect?	29
The Correctness of Affect Assessments.....	30
Filling the Gap.....	30
<i>Relevance and Implications: Why This Study Matters.....</i>	31
<i>Study Objectives</i>	31
Instrument Development Phase.....	31
Prediction Testing Phase.....	32
Methods	32
<i>Instrument Development Phase: Equine Affect Assessment Accuracy</i>	
<i>Indicator</i>	33
<i>Operant Definitions.....</i>	33

ACCURACY OF HORSE AFFECT ASSESSMENTS

Accuracy	33
Terms	34
Categories	34
Correct Categories	34
Incorrect Categories/Terms.....	34
<i>Participants: The Experts</i>	35
Population of Interest and Sample	35
Exclusion and Inclusion Criteria.....	35
Sampling Procedure	35
<i>EAAAI Development Study Design</i>	36
Independent and Dependent Variables	36
How Data Was Measured and Recorded	36
Stimuli: The Videos	36
<i>Procedure</i>	37
<i>Prediction Testing Phase</i>	38
Participants: The Three Groups	38
<i>Prediction Testing Phase Design</i>	40
Independent and Dependent Variables	40
How Data Was Measured and Recorded	40
<i>Procedure</i>	40
Instructions.....	41
Changes to the Prediction Testing Phase’s Activity	41
Results	42

ACCURACY OF HORSE AFFECT ASSESSMENTS

<i>Refining the Experts Survey into the Equine Affect Assessment Accuracy</i>	
Indicator	42
Preparing the Data.....	42
Qualitative Behavioural Assessment	43
Fixed Lists.....	44
Free Choice Profiling.....	44
Decision to use Free Choice Profiling.....	45
Visual Analogue Scale.....	45
Generalized Procrustes Analysis.....	46
The Use of The QBA in Research	46
The QBA's Purpose.....	47
Validity and Application	47
Generalized Procrustes Analysis: Residuals by Object	48
Video Selection Informed by the Generalized Procrustes Analysis	50
Code Development.....	51
The Equine Affect Assessment Accuracy Indicator	52
Prediction Testing Phase	52
Planned Statistical Analysis	52
Preparing the Data.....	53
Composition of the Three Group Samples	54
Correlation of Self-rated Horse Experience and Horse Experience Scores	55
Findings.....	56
Between Subjects One-Way Welch's ANOVA: Main Effect of Profession	57

ACCURACY OF HORSE AFFECT ASSESSMENTS

Findings.....57

One-Way ANOVA on Self-Rated Horse Experience Levels..... 59

Findings.....60

Horse Experience Scores versus Self-rated Levels of Horse Experience... 61

Multiple Regression Analysis 62

Findings.....62

Accuracy of Equine Affect Assessments 64

Total Assessment Accuracy Scores64

How Well do the Experts Perform?64

Discussion66

The EAAAI Worked..... 66

Just How Accurate Were the Groups? 67

Online Delivery..... 68

Footage as Stimuli 68

Are EAMHPs More Accurate than NEAMHPs and Laypeople? 69

Implications of the Findings of the Welch’s ANOVA69

 Regarding NEAMHPs with High Levels of Horse Experience ...69

 Summary.....69

Implications of the Findings of the Multiple Regression Analysis..... 70

Accuracy Improves as Horse Experience Increases..... 70

Importance of the Horse Experience Questionnaire71

The Impact of Horse Experience on Assessment Accuracy71

 Possible Differences in Types of Horse Interaction.....72

ACCURACY OF HORSE AFFECT ASSESSMENTS

The State of Equine Welfare in Equine Assisted Therapy Settings..... 73

Anthropomorphic Descriptions of Affect..... 75

Sample Gender Distribution 77

Limitations..... 77

 Categorization of Terms77

 Possible Impact of Self-Selection Among Laypeople78

 Challenges Related to Participant Recruitment78

 Timeline for Equine Assisted Therapy Involvement When Screening for
 EAMHPs79

Recommendations 79

 Considerations for the Training of EAMHPs79

 Anchoring Accuracy with New Experts80

 Reducing Classification Bias with a Hybrid Term Generation
 Methodology80

 Regarding Self-Rated Levels of Horse Experience81

 Purposes the EAAAI Could be Adapted For82

Conclusion**83**

References.....**84**

ACCURACY OF HORSE AFFECT ASSESSMENTS

List of Tables

Table 1: Instrument Development Phase: Results of the Generalized Procrustes Analysis: Residuals by Object _____	49
Table 2: Descriptive Analysis of the Prediction Testing Phase Participant Sample _____	54
Table 3: Correlations Related to Horse Experience Scores and Self-rated Levels of Horse Experience _____	56
Table 4: Average Total Assessment Accuracy Scores Across All Professional Groups _____	58
Table 5: Average Total Assessment Accuracy Scores Across All Self-Rated Horse Experience Groups _____	60
Table 6: Results of Multiple Regression Analysis Pertaining to Horse Experience Scores and Profession _____	63
Table 7: Average Total Assessment Accuracy Scores Across All Professional Groups including Experts _____	65

List of Figures

Figure 1: Generalized Procrustes Analysis: Videos Organized by Residuals _____ 50

List of Abbreviations

Abbreviation	Definition
EAT	Equine assisted therapy
EAMHP(s)	Equine assisted mental health professional(s)
NEAMHP(s)	Non-equine assisted mental health professional(s)
EAGALA	Equine Assisted Growth and Learning Association
PATH Intl.	Professional Association of Therapeutic Horsemanship International
SPCA	Society for the Prevention of Cruelty to Animals
QBA	Qualitative Behavioural Assessment
FCP	Free choice profiling
VAS	Visual Analogue Scale
REB	Research Ethics Board
SSHRC	Social Sciences and Humanities Research Council
CGS M	Canada Graduate Scholarship-Master
ANOVA	Analysis of variance
<i>M</i>	Mean
<i>SD</i>	Standard deviation
<i>SEM</i>	Standard error of measurement; standard error of mean
<i>95% CI</i>	95% Confidence Interval
<i>N</i>	Total number of cases
<i>n</i>	Subsample number of cases
<i>F</i>	F distribution
<i>p</i>	Probability
τ_b	Kendall's rank ordered correlation coefficient
η^2	Eta-squared; measure of strength of relationship

**Assessment Accuracy of Horse Affect: A Comparison of Equine Assisted
Mental Health Professionals, Non-Equine Assisted Mental Health Professionals, and
Laypeople**

It takes nothing away from a human to be kind to an animal.

—Joaquin Pheonix

There exist many accounts of animal assisted therapy, ranging from anecdotal reports, to session transcriptions, to scientific research. The therapeutic power of animals is discussed in the literature as both mystical and at the same time grounded in reality. Animal assisted therapy has come a long way since its beginnings as little more than anecdotal theorizing by Freud regarding the therapeutic presence his Chow Chow Jofi instilled upon his patients. Yet over the years the field has gained not only public approval, but academic acceptance as well (Coffey, 2013; Shubert, 2012; VanFleet et al., 2019).

While dogs are often the first species to come to mind when people hear the words “therapy animal,” horses have become not only exceptionally popular but indeed a cornerstone species in the practice. This should come as no surprise, given that they are incredibly social animals, and react and respond rapidly through a vast quantity of behaviours (Merkies et al., 2018; Waite & Bourke, 2013). Further, horses seem to possess qualities not dissimilar to the Rogerian core conditions: they are nonjudgmental of humans and do not seem to discriminate based on whether a person is neurotypical or not (Merkies et al., 2018); they display clear and reactive behavioural responses to interactions with clients that are seemingly congruent and honest (Waite & Bourke,

ACCURACY OF HORSE AFFECT ASSESSMENTS

2013); and they are capable of recognizing humans who lack certain knowledge and adjusting their behaviours accordingly (by increasing the frequency and intensity of signalling behaviours; Ringhofer & Yamamoto, 2017). Further, the analogy of a mirror is often used when describing horses, suggesting that through their behavioural reactions, clients are able to self-reflect in ways that foster therapeutic change (Waite & Bourke, 2013).

It is clear from the research that horses represent a good choice for an animal co-therapist, to use a term coined by Dr. Boris Levison, affectionately titled the “father of animal-assisted therapy” (Fine & Beck, 2019, p. 5). However, any time an animal is used for human purposes, their welfare must be protected. The Five Freedoms paradigm has been used to assess and help protect the welfare of animals while drawing distinctions between the physical and mental domains (Mellor & Beausoleil, 2015; Mellor, 2016). It is this mental domain that is most applicable for the purposes of this paper. The associated Five Domains model also possess three orientations, the most relevant of which is the affective state orientation which pays particular attention to the accurate assessment of the motivations, preferences, and aversions of animals, and their related behaviours (Mellor, 2016).

As self-appointed caregivers of the domesticated horse, the decisions humans make and the experiences they bring to the relationship can directly impact the welfare of the equine (Mellor, 2016). Unfortunately, misinterpretations of equine behaviour have been identified as a context in which horse welfare is likely to be jeopardized (DuBois et al., 2018; Horseman et al., 2016). People with little to no equine experience are at greater risk of misinterpreting a horse’s behaviour and affective states, which, if they are not

ACCURACY OF HORSE AFFECT ASSESSMENTS

mentored by a knowledgeable person, may result in the development of undesirable equine behaviours or human injury (Gronqvist et al., 2016). Unfortunately, the development of undesirable behaviours is the leading reason horses between the ages of 2 to 7 are sent to slaughter (Minero & Canali, 2009). Thus, a person's knowledge of, practical experience level with, and ability to accurately interpret the behaviours and affective states of horses, all play an important role in the preservation of equine welfare (Minero & Canali, 2009; see also Hal et al., 2008; Hausberger et al., 2011; Horseman et al., 2016; McGreevy & McLean, 2009; Visser et al., 2009).

Equine assisted mental health professionals (EAMHPs) who are unable to accurately assess the horse co-therapists mental state may misinterpret the animal's behaviour, leading to a breakdown in horse-human communication, and increasing the chances of the equine being placed in situations that jeopardize its wellbeing (Gronqvist et al., 2016; see also DuBois et al., 2018; Horseman et al., 2016; McBride & Long, 2001; Merkies et al., 2018). Therefore, the EAMHP's accuracy at assessing the horse's affective state directly impacts its welfare (Merkies et al., 2018). Since EAMHPs employ horses in their work, they are responsible for the ethical use of the animal and ensuring its mental welfare is not being compromised—as dictated by the Five Freedoms paradigm—while also maintaining the safety of both humans and equines (Merkies et al., 2018). Thus, it is important that these professionals are able to accurately interpret their therapy horse's affective states (Merkies et al., 2018; see also DuBois et al., 2018; Horseman et al., 2016; Mellor, 2016).

Currently, there are no regulations in Canada or the United States that supervise the use of horses in therapeutic contexts. Further, no research to date has investigated

ACCURACY OF HORSE AFFECT ASSESSMENTS

whether EAMHPs can accurately assess the affective states of their horse co-therapists.

This study aimed to do so using a two phased approach.

The objectives of the instrument development phase were to (a) establish a set of stimuli (video recordings) for which the affective states of the horses in them were agreed upon by a sample of expert participants in horse behaviour and (b) to develop both answer keys for the videos representing the correct affective states of those horses as identified by the experts and a code through which to score participants in the prediction testing phase. From this phase the Equine Affect Assessment Accuracy Indicator (EAAAI) was developed, which could be used to determine the accuracy of people's assessments of equine affect based on their observations of the videos.

The objective of the prediction testing phase was to use the EAAAI to investigate whether or not assessment accuracy of horse affect was effected by level of horse experience and/or profession (wherein three groups were assessed: registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices (1) involve horses; (2) don't involve horse, and; (3) laypersons who do not fit into the other two groups).

Literature Review

A Historical Overview of Animal Assisted Therapy

The modern definition of animal assisted therapy is any therapy in which an animal is employed and plays a key role in the therapeutic process (Coren, 2015). Much of the history of animal assisted therapy lies in an anecdotal vein, with varied accounts of the healing or therapeutic qualities of animals. While this information is hardly peer

ACCURACY OF HORSE AFFECT ASSESSMENTS

reviewed, it is nonetheless useful in laying a foundation for a discussion of the practice (Fine & Beck, 2019).

The Healing Properties of Animals

To go back over 4000 years, ancient Egyptians worshipped the god of Death, Anubis, who had the head of a dog (Coren, 2015; Serpell, 2019). Early Egyptians believed that the lick of a dog had the ability to cure ailments, wounds, and disease—a belief later adopted by the Greeks (Coren, 2015; Serpell, 2019). This belief persists today in some cultures; for example, the French possess a modern saying: “A dog’s tongue is a doctor’s tongue” (Coren, 2015, p. xix).

Horses, the subject of this paper, have been part of the human landscape for nearly 6,000 years (Dawkins & Wong, 2004) and acknowledgement of their therapeutic value has deep roots, dating back to ancient Greece, where the horse was understood to provide physical and psychological benefits (Snider et al., 2007). Hippocrates himself believed experiencing the pace and gait of a horse was in and of itself a therapeutic exercise (Granados & Agís, 2011).

The Emergence of Animal Therapy

Like many other forms of scientific inquiry in the Western world, animal assisted therapy was put on hiatus during the Middle Ages; and so it was not until the 17th century that John Locke first suggested that animals might offer therapeutic mental benefits to their human companions. And, not until the 19th century did Florence Nightingale extended this idea, suggesting that animals might help children develop empathy and social skills, and alleviate depression in patients (Serpell, 2019).

ACCURACY OF HORSE AFFECT ASSESSMENTS

York Retreat. One early well documented example of animal use in conjunction with human therapy is thought to have occurred in the 18th century in York, England, at the York Retreat founded by William Tuke (Serpell, 2019; Shubert, 2012). In 1796 Tuke opened the retreat to provide therapy patients with an environment sharply different from that of common institutions at the time (Serpell, 2019; Shubert, 2012). Patients were not required to wear uniforms, were encouraged to engage in meaningful activities (such as reading, writing, and art) and most notably, were free to wander the well-kept grounds and interact with the plethora of animals they found there (Serpell, 2019; Shubert, 2012).

Corporal Wayne and Smoky. During World War II an injured American Corporal by the name of William Wayne, who was receiving care at a Philippine Army Hospital, was visited by his Yorkshire Terrier, Smoky, (Coren, 2015). The little dog's moralizing effect spread throughout the injured soldiers, noticeably raising their moods (Coren, 2015). Having witnessed the morale boosting effect of Smoky firsthand, Corporal Wayne began taking the little dog with him to other hospitals and continued to do so well after the war had ended (Coren, 2015). In our modern day, dogs (and other therapy animals) are still welcomed into hospitals, old-age homes, and hospices, where they instill a therapeutic effect (Coren, 2015).

Freud and Jofi. Freud, too, eventually saw the value of utilizing an animal companion in the treatment of patients (Shubert, 2012). While at first Jofi's presence was intended to comfort Freud himself, he soon came to believe that the presence of the amiable Chow Chow also put his clients at ease, encouraging feelings of security and acceptance which aided in their therapy (Coren, 2015; Shubert, 2012). Freud noted this calming effect was evidenced further by the decreased intensity of client resistance as

ACCURACY OF HORSE AFFECT ASSESSMENTS

they neared grasping the source of their problems, as compared to patients who were not treated in the presence of the animal (Coren, 2015). Freud's use of Jofi marks just one early instance in which an animal was used as an object of transference for human therapy, and Freud's notes on the effects she had on clients in his sessions have come to aid in the modern-day application of therapy dogs (Coren, 2015).

Levinson and Jingles. Perhaps the most prominent figure in animal assisted therapy was Boris Levinson (Fine & Beck, 2019), who could be said to have discovered the therapeutic effects of dogs by complete accident (Shubert, 2012). As a child psychologist, Levinson had been working with a difficult young boy who was closed off and seemed uninterested in forming the therapeutic bond with Levinson that was the key to his treatment (Coren, 2015; Levinson & Mallon, 1997; Shubert, 2012). One day the boy's mother dropped him off early and they were greeted by a small dog named Jingles (Levinson & Mallon, 1997; Shubert, 2012). That moment changed the course of that little boy's treatment, as he suddenly became open to, and interested in, the therapy by way of Jingles (Levinson & Mallon, 1997; Shubert, 2012). This experience convinced Levinson, beyond a shadow of a doubt, that utilizing an animal as a co-therapist could jumpstart the development of the therapeutic relationship between client and practitioner (Levinson, 1978). Levinson then set out to incorporate animals in his therapeutic practice and present his discovery at conferences, in order to rally interest and intrigue amongst medical health professionals—encouraging them to utilize animals in their own practice even in spite of the mockery and ridicule with which he was greeted (Shubert, 2012).

Animal assisted therapy has come a long way from its humble beginnings being laughed at in conferences; it has notably gained acceptance not only with the public but

ACCURACY OF HORSE AFFECT ASSESSMENTS

amongst academics as well (Coffey, 2013; see also VanFleet et al., 2019). And, while dogs may come to mind as the primary example of a therapy animal, a plethora of other species are also suitable, including, but not limited to: guinea pigs, chickens, goats, sheep, cats, fish, birds, some exotic species, donkeys and of course, horses (Coren, 2019).

Horses in Psychology

Equines, in particular horses, have made their mark on the field of animal assisted therapy, and have become one of the cornerstone species of the practice. In Canada and the USA horses are common, with approximately 500,000 horses and “855,000 people active in the horse industry” in the former (Fédération Equestre Internationale, 2020) and 7.2 million horses in the latter (American Horse Council Foundation, 2018). Representing ratios of 1:80 (Statistics Canada, 2023) and 1:47 (United States Census Bureau, 2023) horses to humans in Canada and the USA respectively.

Therapeutic Traits of Horses

Horses are social animals, which respond to and interact with one another through an array of immediate visible non-verbal, and audible verbal behaviours (Merkies et al., 2018; Waite & Bourke, 2013). They are also large and powerful prey animals, and thus their physical presence demands caution and attentiveness, and requires the development of clear non-verbal communication (Waite & Bourke, 2013). Their therapeutic value lies in the horse’s clear and reactive behavioural responses to interactions with and perceptions of the client (Merkies et al., 2018; Waite & Bourke, 2013). They have also been likened to mirrors, allowing clients to engage in deep, meaningful self-reflection which can lead to therapeutic change and growth (Waite & Bourke, 2013).

ACCURACY OF HORSE AFFECT ASSESSMENTS

How Horses Help Humans

Physically. Considerable research has been conducted on the physical and psychological benefits horses bring to humans. It is believed that people with a variety of conditions benefit from working with a therapeutic horse (Latella & Abrams, 2019). Conditions researched have included but are not limited to: “cerebral palsy, multiple sclerosis, down syndrome, developmental delay, autism, stroke, traumatic brain injury, spinal cord injury, spina bifida, convulsive disorders, amputation, and muscular dystrophy” (Latella & Abrams, 2019, pp. 145–146).

Cognitively. Further, in a table compiled by Latella and Abrams (2019) in Chapter 10 of the *Handbook on Animal Assisted Therapy*, they summarized just a few of the qualitative and quantitative findings garnered from research in the field of equine assisted therapy, which include (but are not limited to): grieving children developing greater self-esteem, self-confidence, and trust; children with autism spectrum disorder developing increases in affective understanding, and initiating and maintaining interactions; adults with balance deficits developing improvements in their balance and quality of life; children with pervasive developmental disorder seeing increases in emotional expression, better adaptation to change, decreases in fixative behaviours, and increases in their ability to communicate. In adult combat veterans, increases in sociability and trust of others, along with reduced isolating tendencies were observed; and youth in one study saw an increase in self-esteem, sense of responsibility, and an increase in core competency.

ACCURACY OF HORSE AFFECT ASSESSMENTS

The Horse as Co-Therapist

Horses are appropriate models for the term of co-therapist, as, like other commonly employed animals, they seem to possess the core Rogerian components: They are non-judgmental (Merkies et al., 2018; Waite & Bourke, 2013), promoting feelings of warmth, engagement, acceptance, and trust (unconditional positive regard; Waite & Bourke, 2013; see also Jenkins et al., 2014). They are responsive to human interactions (Waite & Bourke, 2013), and through client interpretations may be perceived as genuine, honest (i.e., congruent), and empathetic in nature (Jenkins et al., 2014; Lac & Walton, 2012; Merkies et al., 2018; Ringhofer & Yamamoto, 2017; Waite & Bourke, 2013). Horses are often incorporated into therapy programs because of their perceived congruence, as their behaviour is directly reflective of their mental state and interactions with humans (Lac & Walton, 2012; Waite & Bourke, 2013; see also Ringhofer & Yamamoto, 2017). Further, as a social species, horses display uncomplicated responses to interactions with others (Waite & Bourke, 2013). It is believed that the immediate and direct feedback a horse gives through its visible behavioural response provides valuable information to the client regarding their own thoughts and behaviours (Merkies et al., 2018; Waite & Bourke, 2013).

Indeed, research even suggests that horses are intelligent enough to distinguish between humans based on a person's knowledge of events (which the horse seems to intuitively understand) and will choose to alter their behaviour based on this inference (Ringhofer & Yamamoto, 2017). It is clear that horses are powerful, not only physically but also in the psychological and emotional impact they can have on humans, and that their therapeutic abilities are far-reaching.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Types of Equine Assisted Therapy. Horses as therapy animals have become notable to the point that special terminology has been created for models involving them. Therapeutic, riding for example, is a form of therapy that predominantly involves riding the horse; while it focuses on achieving both mental and physical client goals, it is of most benefit to the latter (Bass et al., 2009; Bona, 2020).

In equine facilitated psychotherapy, the therapy session is primarily conducted by two experts: One who handles the animal and is a credentialed equine professional, and another who is the therapist (Bona, 2020; Nilson, 2004). Most notably, this form of therapy requires that the acting psychological professional be a registered psychotherapist; however, this role and that of the equine professional are sometimes filled by the same person (Bona, 2020; Nilson, 2004).

Equine facilitated counselling is not dissimilar to equine facilitated psychotherapy and differs only in that the registered psychotherapist is replaced by a qualified and certified counsellor (Bona, 2020). In both models, talk-therapy is often utilized to a great extent along with ground-based activities (rather than in-saddle activities; Bona, 2020).

On the other hand, equine experiential or assisted learning is unique in that it may be conducted by a facilitator (who may or may not be trained in mental health), rather than a registered counsellor or psychotherapist (Bona, 2020). Also, in addition to using a team of four approach (comprised of the horse, client, facilitator, and horse expert), the client may be either an individual or a group (Bona, 2020).

Associations. The Equine Assisted Growth and Learning Association (EAGALA) model involves a therapist working alongside an experienced equine specialist (where these two roles are never filled by the same person) in a team of four, with the horse and

ACCURACY OF HORSE AFFECT ASSESSMENTS

client (EAGALA, n.d.). Unlike most of the previously mentioned forms of equine assisted therapy, EAGALA takes an exclusively ground-based approach in which riding never occurs (EAGALA, n.d.).

Lastly, the Professional Association of Therapeutic Horsemanship International (PATH Intl.) model supports its members in the provision of therapeutic/adaptive horsemanship, equine-assisted learning, equine-assisted services for veterans, and incorporating equines into counselling/therapy (PATH Intl., n.d.-a). Members with or without mental health professional credentials can become PATH Intl. certified after completion of the required prerequisites. Both EAGALA and PATH Intl. promote the ethical use of horses in therapy.

Summary. Here-within, equine assisted therapy will be the umbrella term used to denote any of the above-mentioned models of therapy in which an equine (horse, donkey, or mule) is actively involved in a client's therapeutic treatment and goals. For further details about these equine assisted therapy models, please refer to Appendix A.

Concerns of Equine Welfare

In this section, equine welfare related research, topics, and definitions will be introduced and discussed. The protection of equine welfare and the ethical use of horses in therapeutic settings are the foundational values for this thesis and therefore their discussion is critical to understanding the context within which this research study is situated.

Defining Affect

The word *affects* means “a set of observable manifestations of an experienced emotion: the facial expressions, gestures, postures, vocal intonations, etc., that typically

ACCURACY OF HORSE AFFECT ASSESSMENTS

accompany an emotion” and “the conscious emotion that occurs in reaction to a thought or experience” (Merriam-Webster, n.d.-a, see Noun section). The term affect has commonly been used when discussing animal welfare in regard to the Five Freedoms paradigm and the Five Domains model. Mellor (2016) proposed several characteristics of *affect* or affective states: First, both negative and positive affective states exist.

Negative Affect. Negative affective states are the result of disruptions to the physical or functional state of the animal’s body or cognition (Mellor, 2016). Negative affective states can be internally or externally generated, in which the former refers to affects such as breathlessness, thirst, or hunger, which motivate the animal to fulfil its physical needs (Mellor, 2016). Externally generated negative affects, on the other hand, encompass emotional terms such as anxiety, fear, anger, boredom, and depression which arise as a result of the conditions and circumstances the animal finds itself in (Mellor, 2016).

Positive Affect. Positive affective states result from an animal’s mental perception of the amiable external factors of its situation and environment (Mellor, 2016). They are experienced when the animal is given the chance to engage in behaviours which are rewarding to them, such as forming bonds with conspecifics or being kept in a safe environment (Mellor, 2016).

A Brief Overview of The Five Freedoms Paradigm

Whenever an animal is used for human purposes, even for animal assisted therapy, its welfare must be taken into consideration and protected. The Five Freedoms paradigm was originally developed in 1965 and formalized in 1979 as a guideline for the appropriate confinement of animals (Farm Animal Welfare Council, 2012). Later it was

ACCURACY OF HORSE AFFECT ASSESSMENTS

extended by John Webster in 1993/1994 (Webster, 1994, 2008) to be a tool for the assessment of animal husbandry protocols and has since remained unaltered (Mellor, 2016). Today, it is internationally recognized and utilized for this purpose and has contributed greatly to the “development of evidence-based animal welfare standards globally” (SPCA, n.d.; see also Mellor, 2016).

The Five Freedoms and their accompanying provisions are:

1. Freedom from thirst, hunger, and malnutrition:
 - 1.1. By providing ready access to fresh water and a diet to maintain full health and vigour.
2. Freedom from discomfort and exposure:
 - 2.1. By providing an appropriate environment including shelter and a comfortable resting area.
3. Freedom from pain, injury, and disease:
 - 3.1. By prevention or rapid diagnosis and treatment.
4. Freedom to express normal behaviour:
 - 4.1. By providing sufficient space, proper facilities, and company of the animal’s own kind.
5. Freedom from fear and distress:
 - 5.1. By ensuring conditions and treatment which avoid mental suffering (Webster, 1994, 2005, as compiled by Mellor, 2016).

While the words “freedom from” are used, this is clearly understood to indicate that an animal should be as free as possible from these negative physical and mental

ACCURACY OF HORSE AFFECT ASSESSMENTS

states, and that it is impossible for an animal to be completely free of any of them permanently (Mellor, 2016; Mellor & Beausoleil, 2015).

Physical Domains. The Five Freedoms paradigm makes distinctions between four physical domains and a fifth mental domain (Mellor & Beausoleil, 2015). The first four freedoms (relating to physical domains), reflect that negative affective states are the result of disruptions to the homeostasis of the animal's internal physical or functional state (for example, lack of water leads to thirst and dizziness/listlessness; Mellor, 2016).

Mental Domain. The last freedom, however, reflects the mental domain. Regarding the mental domain, negative affective states are said to result from sensory inputs which contribute to an animals' cognitive assessments of its environment and current situation (for example, a lack of conspecifics may result in loneliness and depression; Mellor, 2016).

A Brief Overview of The Five Domains Model

The sentience of animals and recognition of their emotional needs led to the development of the notions: "a life not worth living," "a life worth living," and "a good life" by the UK's Farm Animal Welfare Council (2009, p. iii). In response to this shift, the Five Domains model, developed by Mellor and Reid (1994) grew out of the Five Freedoms paradigm to emphasize "a greater focus on the mental state of an animal (how the animals 'feels') and acknowledges [*sic*] that welfare can be both positive and negative" (SPCA, n.d. para. 4; see also Farm Animal Welfare Council, 2009).

There exist within the Five Domains Model three orientations:

1. Biological function,
2. Natural living, and

ACCURACY OF HORSE AFFECT ASSESSMENTS

3. Affective state (Mellor, 2016).

The affective state orientation has gained credibility as the research regarding animal subjective experience has grown, and particular attention has been paid to the identification and interpretation of an “animal’s motivations, preferences, aversion, and aligned behaviours” (Mellor, 2016, p. 14). In fact, sentient animals, including horses, are likely much more keenly aware of their environmental conditions and social factors than has previously been thought (Mellor, 2014). In light of this, Mellor (2014) proposed an extension beyond sole consideration for limiting negative affective states (as seen in the Five Freedoms Paradigm), towards the inclusion of positive ones as well (as seen in the Five Domains Model; Mellor, 2014). By providing the animal with opportunities to engage in behaviours it finds pleasurable and pleasant, the animal’s welfare is significantly enhanced (Mellor, 2014, 2016).

As the caregiver in the horse-human relationship, our behaviours and decisions directly influence the welfare of the equine. As such, EAMHPs must be capable of accurately assessing their co-therapist’s emotions so that they can not only negate influences that may lead to negative affective states, but also provide the horse with opportunities to experience positive affective states as well (Mellor, 2016).

Contributors to Equine Welfare Risk

In an archival analysis of the current literature on equine welfare, Minero and Canali (2009) summarized areas of concern and their implications with regards to horses. Improper husbandry, poor animal training skills, and lack of knowledge have all been found to contribute to the development of dangerous and undesirable behaviours in horses

ACCURACY OF HORSE AFFECT ASSESSMENTS

(Horseman et al., 2016; Minero & Canali, 2009), such as rearing, biting, kicking, and stereotypies¹.

Training and Riding. The process of training and riding a horse was found to potentially contribute to the development of undesirable or dangerous behaviours and poses a context in which equine welfare may be compromised (Minero & Canali, 2009). Horses may be put in mentally or physically harmful situations if the handler overestimates the horse's cognitive and reasoning abilities (Minero & Canali, 2009) or lacks knowledge of learning theory and thus attempts to train the horse using inconsistent signals, and harsh and/or non-contingent punishments (Minero & Canali, 2009). The inappropriate use of the horse (e.g., asking it to do something it is physically incapable of doing), employing punishment-based training, and pushing a horse to work harder than it should, have been identified as likely contributors to compromised equine welfare (DuBois et al., 2018; Horseman et al., 2016).

Training History. It is critical to consider that horses are often gifted to equine assisted therapy programs or adopted into them and thus the training history of the animal may be unknown. Horses who have been subjected to punishment-based methods of training may exhibit stunted behavioural expression (due to learned helplessness, a psychological condition which results from a critical loss of control over one's

¹ Stereotypies evolve out of inappropriate husbandry and housing and are repetitious, seemingly purposeless behaviours almost akin to human compulsions (McGreevy, 2004; Waters et al., 2010) which “enables stressed animals to cope with stress” (Minero & Canali, 2009, p. 224). Stereotypies are often accompanied by other unwanted behaviours such as: stall-kicking, head-tossing, pawing, wood-chewing, and licking (to name a select few, for a more thorough list see McGreevy, 2004).

ACCURACY OF HORSE AFFECT ASSESSMENTS

environment; Webster, 1994; see also Hal et al., 2008) which may make it difficult for inexperienced professionals to accurately assess their affective state.

Lack of Knowledge. Canadian equine industry members agree that there are welfare issues facing horses in the industry, and that inappropriately trained horses and horses with unknowledgeable owners or caregivers are particularly at risk of poor welfare (DuBois et al., 2018). In fact, owner ignorance and lack of knowledge were believed to be the most frequent cause “of reduced horse welfare” (DuBois et al., 2018, p. 7).

These findings were also observed in a survey of equine stakeholders in Great Britain (Horseman et al., 2016). Of particular interest, the stakeholders agreed that horse welfare was likely to be compromised when humans misunderstood the horse’s behaviour, especially when stress or pain went uncorrected (Horseman et al., 2016). Such factors can often lead to the development of unwanted behaviours (another context the stakeholders identified in which horse welfare may be compromised; DuBois et al., 2018; Horseman et al., 2016)). In turn, such behaviours may lead to the sale or auction of the horse, which can pose even greater welfare infringement issues (DuBois et al., 2018; Horseman et al., 2016).

Consequences. Thus, training and the accurate interpretation of horse behaviour and affect play an important role in equine welfare (Gronqvist et al., 2017). Overall failure to train a horse often results in behavioural problems (Minero & Canali, 2009; Horseman et al., 2016) which are identified as the main reason 66.4% of horses between the age of 2 and 7 are sent to slaughter (Minero & Canali, 2009). Therefore, inappropriate husbandry, training, and lack of knowledge not only negatively impact a horse’s mental and physical wellbeing but also greatly increases the animal’s risk of being slaughtered

ACCURACY OF HORSE AFFECT ASSESSMENTS

(Minero & Canali, 2009). For more details regarding the equine welfare risk contributions of improper husbandry, training, and knowledge, please see Appendix B.

Equine Welfare in EAT

The general belief in equine assisted therapy programs appears to be that horses interpret and respond to human emotions, thus acting as catalysts of change, with some practitioners championing the horse as capable of intuitively knowing what kind of emotional support a client needs (Merkies et al., 2018, see page 62). Unfortunately, this type of anthropomorphizing and placing of unrealistic expectations on the horse may lead to welfare issues, as has been presented.

The EAGALA model's (see Appendix A) emphasis on upholding a high ethical standard for the use of horses in therapy offers a benchmark of hope. This concern, not only for human welfare, but also for that of the horse, emphasizes a valuable consideration for the practice of equine assisted therapy.

Attitudes and Beliefs of Equine Assisted Professionals. In the survey portion of a mixed methods senior honours thesis study conducted by Reega (2017), 59 current certified equine assisted activities and therapy professionals were asked to answer questions intended to explore their current views and experiences related to the welfare of and stress placed on equines involved in EAT. Overall, the answers provided by the survey participants indicated an awareness of potential welfare issues that may arise for the equine, along with a trend to value the sharing and continued gathering of knowledge regarding equine welfare (Reega, 2017). Presented here are some of the most relevant results from the Reega (2017) study with regards to ethical considerations and concerns in EAT.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Education is Important. Interestingly, 80% of the participants indicated that they believed it was extremely important for them to continue their own education regarding equine behaviour and equine welfare (Reega, 2017). In addition, 85% indicated that they felt it was either extremely important or very important to inform “participants/clients about equine behaviour and welfare prior to working with an equine” (Reega, 2017, p. 11). Research by Merkies et al., (2018) compliments these findings, suggesting that equine assisted therapy programs hinge on the facilitator being highly experienced at accurately interpreting horse behaviour to maintain a positive and safe environment that fosters therapeutic change.

Therapy Can be Unsafe for Horses. Further, the participants of the Reega (2017) study also demonstrated they could (and indeed had) recognize situations in which the therapy animal’s safety or welfare was being compromised. Forty-nine out of the 59 respondents indicated they had “chosen to remove an equine from a session in progress due to concerns about the safety and/or wellbeing of the equine” (Reega, 2017, p. 12). Of note, participants tended to strongly or somewhat agree (12% and 40% respectively), with the statement “I have experienced a participant trying to manipulate, scare, or tease an equine in an [EAT] session which I’ve facilitated or observed” (Reega, 2017, p. 13). While the phrasing of the question makes it clear the professional was not the one engaging in the inappropriate behaviour with the animal, this does raise questions regarding the treatment of equine co-therapists by clients in therapy.

Horses Mostly Enjoy Being Co-Therapists. The survey further included three questions directly pertaining to equine wellbeing in sessions. Responses demonstrated that only one participant believed that horses “always” enjoy being involved in EAT, with

ACCURACY OF HORSE AFFECT ASSESSMENTS

the majority of participants (36) responding, “most of the time,” thirteen participants selecting “about half of the time,” and nine selecting “sometimes” (Reega, 2017, p. 14). These results are not necessarily surprising, and perhaps reflect debate over the use of the word “enjoy” in the question, or simply reflect a non-ego centric view of equine benefit in human therapy programs. Regarding the latter, 47% of respondents indicated equine assisted therapy sessions are mutually beneficial for both the equine and human “most of the time” (p.14).

EAT is Somewhat Adequately Regulated. When presented with the statement “in some situations, [EAT] may be detrimental to the well-being of an equine,” 30% of participants strongly agreed, 47% indicated they somewhat agreed, 11% neither agreed nor disagreed, and only 8% somewhat disagreed (Reega, 2017, p. 14). When asked if there “are currently enough rules, guidelines, and regulations to ensure the well-being of the equine involved in an [EAT] session” only 2 participants strongly agreed, while the majority (26 participants) somewhat agreed (Reega, 2017, p. 14). A similar pattern of agreement was found when participants were asked if “concerns related to equine welfare in [EAT] are being appropriately considered and addressed,” to which 3 individuals strongly agreed and the majority (24 participants) somewhat agreed (Reega, 2017, p. 15).

Summary. It is clear from these responses that professionals involved with equine assisted therapy are aware of instances in which equine welfare may have been compromised or nearly so during a therapy session and believe that an equine’s involvement in therapy is not always beneficial to the animal (Reega, 2017). However, the results also suggest participants believe that progress can still be made regarding the

ACCURACY OF HORSE AFFECT ASSESSMENTS

implementation of rules, guidelines, and regulations in equine assisted therapy and that the process of handling and addressing concerns can still be improved (Reega, 2017).

Impacts of Client Populations on Equines

In therapy, horses may often not be given a choice about which clients they work with. Since the majority of therapists interviewed in Reega (2017) noted they had seen a horse being inappropriately interacted with during a session, it is worth considering briefly whether specific client populations pose any increased risk to equine co-therapists.

Youth. Kaiser et al. (2006) found that being ridden by at-risk youth caused therapy horses to experience significantly more stress compared to when they were being ridden by youth who had (a) no handicaps, (b) who were physically handicapped, (c) who were psychologically handicapped, or (d) special education children. Further, at-risk youth were more likely to act out towards the horse and see the animal as a tool (Kaiser et al., 2006).

However, the authors expressed that at-risk youth should not be excluded from therapeutic riding programs (Kaiser et al., 2006). Instead, to avoid compromising equine welfare, special management and workload considerations should be implemented for any therapy horse working with at-risk clients (Kaiser et al., 2006).

Experienced Riders. Notably, being ridden by an experienced rider for training or re-training purposes was found to be significantly more stressful for a horse than therapeutic riding, regardless of the nature of the client (Kaiser et al., 2006). Therefore, therapeutic riding should not be considered as any more stressful to horses than recreational or non-therapeutic riding (Kaiser et al., 2006).

ACCURACY OF HORSE AFFECT ASSESSMENTS

Clients with PTSD. Merkies et al. (2018) found that horses, when working with a person diagnosed with PTSD versus a neurotypical person, did not demonstrate any alterations in their behaviour (Merkies et al., 2018). In this study, the neurotypical individuals were recruited based on their physical appearances so that they closely resembled the PTSD participants they were paired with (Merkies et al., 2018).

Furthermore, the neurotypical individuals intentionally mirrored the movements and behaviours of their PTSD counterpart during their session with the horse (Merkies et al., 2018). Salivary cortisol, collected from the horses to assess stress levels pre- and post-session, revealed no significant increases, indicating that neither client type presented any unique stress to the animal (Merkies et al., 2018). Interestingly, horse heart rate did differ between client types. Horse heart rates increased steadily when they worked with PTSD individuals, and their heart rates continued to climb even after the session had ended and the PTSD individual had left the pen (Merkies et al., 2018). In comparison, horse heart rates decreased in the presence of neurotypical participants and remained constant once the session had ended (Merkies et al., 2018). These results perhaps indicate that while the horse experienced a difference between the participant types (or was perhaps aware of the difference), it did not feel significantly stressed by either type of interaction, nor did it feel the need to alter its behavioural responses.

Summary. These findings may offer support to the notion that therapeutic work is not uniquely stressful to the equine co-therapist. The findings may also support a benefit to using an equine in therapy, as their lack of behavioural discrimination between neurotypical and PTSD diagnosed participants may provide a sense of normalcy (Merkies

et al., 2018) and non-judgemental acceptance (Merkies et al., 2018; Waite & Bourke, 2013) for the client.

Making Meaning of Horse Behaviour

We have just discussed how a lack of behavioural discrimination among various client groups can be a beneficial feature of the horse as co-therapist. But in order to know whether or not an animal's welfare is being compromised, the professional still needs to be able to assess the horse's behaviour and meaningfully interpret it, regardless of how or whether it varies between clients.

The Importance of Clear Communication

Gronqvist et al. (2016) noted that most (human) injuries or accidents that occur in equine veterinary practice can be attributed to a breakdown in communication between human and horse. As the cognitively superior species, and the one who is usually manipulating the animal, it is clear that the responsibility to find creative ways to establish and maintain effective communication falls to us. Horses communicate predominantly through non-verbal behaviours and thus, a breakdown in communication between horse and human represents a breakdown in our understanding of their behaviour—understanding that requires knowledge and accurate interpretations of the objectives of the animal's actions and its overall demeanour. As previously discussed, DuBois et al. (2018) and Horseman et al. (2016) identified misinterpretation of behaviour to be a potential context in which horse welfare is likely to be compromised, thus it is integral that we correctly assess a horse's affective state and the intentions behind their behaviour (DuBois et al., 2018; Horseman et al., 2016).

ACCURACY OF HORSE AFFECT ASSESSMENTS

Horses Trying to Communicate with Humans. Interestingly, horses have been found to attempt to send signals to humans when faced with unsolvable tasks, which indicates that horse behaviour is meaningful and important in horse-human communication (Ringhofer & Yamamoto, 2017). When food was hidden in a bucket by one human outside of the horse's pen, the animal would attempt to communicate to caretakers through visual and tactile signals (Ringhofer & Yamamoto, 2017). Furthermore, these behaviours significantly increased when the horse was attempting to communicate to a naïve human, indicating horses may be able to differentiate between humans based on their states of knowing (Ringhofer & Yamamoto, 2017). These findings, presented here and elsewhere in the research, reinforce the importance of attending to, and accurately assessing the behaviour of horses to understand their mental and affective states as they go out of their way to try and communicate with us (Ringhofer & Yamamoto, 2017).

Implications for Equine Assisted Professionals. However, such horsemanship takes time and experience to master; there exist no shortcuts and it cannot be learned in mere minutes or hours (Gronqvist et al., 2016). It stands to reason that therapists involved in equine assisted therapy programs spend a great deal of time around horses (although they may be accompanied by a horse expert in some models) and may expose potentially unknowledgeable clients to them. As such, EAMHPs should possess a medium or (preferably) high level of horsemanship skill to maintain a safe environment for all.

The Influence of Horse Experience

If EAMHPs represent a group expected to be proficient in horsemanship, then it is worth considering how this level of expertise influences their ability to accurately

ACCURACY OF HORSE AFFECT ASSESSMENTS

assess the therapy horse's affective state. In their study, Gronqvist et al. (2016) investigated the unique challenges horses pose to the training of veterinary students and found that less than half of the students in the study were able to correctly assess the affective state of the horse.

Lack of Experience Leads to Misinterpretations. It was proposed by these authors that individuals with low/no horse experience are more likely to be injured because their lack of experience hinders their ability to interpret and anticipate the behaviours and intentions of the horse (Gronqvist et al., 2016). The implication then is that horse-experienced individuals should be better able to meaningfully interpret the intentions and affective states of the horse (Gronqvist et al., 2016). In so doing, horse experienced individuals should be more capable of fully understanding and anticipating a horse's behaviours, allowing the person to predict what the animal will do next and safely respond (Gronqvist et al., 2016).

How A Human's Horse Experience Impacts the Animal. Horses have also been found to behave differently depending on the experience level of the human they are interacting with (Merkies et al., 2018). Overall, horses prefer the company of people to being alone, and horses approached faster and stood closer to an experienced human following a join-up routine² (Merkies et al., 2018). When the human was clear in their

² Join-up refers to a traditional technique involving chasing/urging the horse to canter or gallop around the circumference of a round pen until they display a certain repertoire of behaviours, at which point the negative stimuli is entirely removed, and remains terminated so long as the horse leaves the fence to approach the human (Krüger, 2007). If the horse does not leave the fence within an arbitrary time limit, they are chased again (Krüger, 2007).

ACCURACY OF HORSE AFFECT ASSESSMENTS

body language and cues (as associated with having experience) the horses also displayed more definitive responses (Merkies et al., 2018).

Interestingly, horses in the presence of an experienced human displayed an increased heart rate, while in contrast, horses in the presence of someone who had self-rated fearfulness of horses and therefore likely were inexperienced, displayed lower heart rates (Merkies et al., 2018). The explanation for these findings may lie in the more direct, confident, and potentially predatory behaviour of experienced humans which may be perceived as requiring more attentiveness on the part of the horse, as compared to the possibly softer, clumsier, less intense behaviour of an inexperienced individual.

Summary. Based on the research of Merkies et al. (2018) horses respond differently both physiologically and behaviourally depending on whether the person present has horse experience or not. Unsurprisingly, Merkies et al. (2018) found that when individuals delivered clear cues (as associated with having horse experience) the horse demonstrated more precise and definitive responses.

Significance to Equine Assisted Therapy

The studies by DuBois et al. (2018) and Horseman et al. (2016) revealed equine welfare is at risk of compromise when the animal is handled by someone who is inexperienced and frequently misunderstands the horse's behaviour. It is necessary then that everyone involved in the care and handling of the horse be knowledgeable of horse behaviour and be accurate in assessing the animal's affective state (Minero & Canali, 2009). Clear, humane handling of the horse not only benefits the human by enhancing their safety, but also fosters greater performance and better welfare for the animal (Minero & Canali, 2009). It is pertinent then that the equine assisted therapy facilitator be

ACCURACY OF HORSE AFFECT ASSESSMENTS

a horse-experienced individual in order to maintain a safe and positive learning environment during therapy sessions. This stands to benefit not only the client but to also protect the welfare of the equine via correct interpretations of behaviours leading to accurate assessments of the horse's affective state.

Speaking Another Language: Can People Truly Understand Horses?

It is important that those who work with horses, such as equine assisted mental health professionals, be proficient and knowledgeable of equines (Gronqvist et al., 2016; see also Horseman et al., 2016; McBride & Long, 2001; Merkies et al., 2018). Lack of knowledge and, in turn, breakdowns in clear horse-human communication can lead to the development of undesirable behaviours which can threaten the horse's welfare (DuBois et al., 2018; Horseman et al., 2016; Minero & Canali, 2009) and result in human or horse injury (Gronqvist et al., 2016).

What is Clear Communication?

Clear interspecies communication is commonly considered to have occurred when a human is able to use verbal and/or non-verbal cues and/or artificial aids (e.g., a touch target) to produce an outcome wherein the animal does what the human wanted it to do cleanly and without errors or resistance in exchange for desired consequences (Heidenreich, 2012). However, the likelihood of an animal performing the desired behaviour is not only impacted by clear communication, but the animal's own emotional state which can also play an important role (Starling et al., 2013).

How Emotion Impacts Animal Behaviour. The affective states of animals has been shown to impact behavioural outcomes; for example, it was found in dogs and horses that “as arousal increases in the condition of negative affective state, affective state

ACCURACY OF HORSE AFFECT ASSESSMENTS

is likely to become more negative, and the probability of the animal offering a desired behaviour is likely to decrease” (Starling et al., 2013, p. 314) and further “positive affect appears to have benefits for focus and performance in learning new tasks” (p. 313).

Therefore, to clearly communicate with animals, their affective states should be considered. Failure to correctly interpret the animal’s affective state and adjust/respond to it accordingly may not only decrease the likelihood of a desired behaviour being offered (Starling et al., 2013) but may also lead to or be reflective of ongoing risks to the animal’s welfare (DuBois et al., 2018; Horseman et al., 2016; see also McBride & Long, 2001).

Are People any Good at Assessing Animal Affect?

In one study, humans were found to be able to accurately assess the emotional states of horses based on recorded vocalizations 64% of the time (Merkies et al., 2021). Another study found that humans are more likely to correctly interpret the emotional valence of animal voices (a dog, chimpanzee, and tree shrew) if they are knowledgeable about the species (Scheumann, et al., 2014). However, in both studies, the recordings only presented two affective contexts: positive or negative (Merkies et al., 2021) and affiliative or agonistic (Scheumann, et al., 2014). Interpreting the nuances of the emotional state of another species who cannot correct any misunderstandings is no easy feat (see Starling et al., 2013). And, while veterinarian students in training have been found to struggle to correctly assess the affective states of horses (Gronqvist et al., 2016), surprisingly even those who manage equine facilities are often ignorant of the causal factors, purposes of, and appropriate solutions to stereotypies among horses (which demonstrates lack of affective knowledge; McBride & Long, 2001).

ACCURACY OF HORSE AFFECT ASSESSMENTS

The Correctness of Affect Assessments The idea of *correctly* assessing affect is problematic. Without the extensive and potentially invasive use of measurement devices (for example, the use of a heart rate monitor or measuring horse salivary cortisol from collected samples)—and even then—determining the true emotional state of another species possess challenges. Rather than assessing an animal’s ongoing biometrics to determine emotional state, it would be more practical to develop a non-invasive instrument through which the accuracy of a particular person’s assessments of a species’ (i.e., horses) affective states could be measured.

Filling the Gap. Clearly there is a need for the development of such an instrument and this study sought to do just that. Socially accepted experts on horse behaviour (i.e., equine veterinarians, professional trainers, and credentialed animal behaviourists who work with equines) were selected as the best and most acceptable non-biometric based means for determining the most correct assessments of the affective states of horses observed in video recordings. Video recordings were used so that the stimuli remained consistent between experts and to permit online delivery.

Those videos for which this panel of experts most agreed on the emotional states of the horses could then be retained, along with the terms they used to describe the horses, while those videos for which there was less consensus among the experts could be discarded. In so doing, an instrument for testing assessment accuracy could be developed—whereby the videos could be shown to participants who were then asked to describe the animal, and the resulting terms could then be compared to the answers of the experts to determine the participant’s level of accuracy. Thus, a non-invasive instrument to measure affect assessment accuracy was developed in the first phase of this study

ACCURACY OF HORSE AFFECT ASSESSMENTS

which could then easily be delivered to participants, as was done in the prediction testing phase.

Relevance and Implications: Why This Study Matters

Research has demonstrated that misunderstanding equine behaviours, and therefore inaccurate assessments of affective states, often pose horse welfare concerns. The equine assisted mental health professional is responsible for the ethical use of the horses they employ and ensuring the safety of both humans and equines. To this end, it is important that these professionals are accurate at interpreting the affective states of the horse co-therapist, as this is the only way to ensure the animal's mental welfare is not being compromised. Further, as neither Canada nor the USA have mandatory regulations in place for equine assisted mental health professionals, the findings of this study could be critical for identifying potential equine ethical welfare issues within the field. Lastly, the instrument introduced and utilized in this study could be used in the future development of a competency test for prospective equine assisted therapy professions.

Study Objectives

Instrument Development Phase. In the instrument development phase of this study, expert participants were shown 20 videos of horses and asked to describe the affective states of the animals. The first objective of this phase was to identify and retain 10 videos of the 20 shown that had the greatest degree of consensus amongst the expert participants (i.e., which videos the experts most agreed on the affective states of the horses in them).

The second intention of the instrument development phase was to use the terms generated by all of the experts for each of the preserved 10 videos to form answer keys,

ACCURACY OF HORSE AFFECT ASSESSMENTS

and to develop a code through which the participants of the second (prediction testing) phase could then be scored on the accuracy of their assessments of affect. These 10 retained videos, related answer keys, and scoring codes culminated in the development of the Equine Affect Assessment Accuracy Indicator (EAAAI).

Prediction Testing Phase. The objective of the prediction testing phase was to use the EAAAI to investigate whether or not assessment accuracy of horse affect was effected by level of horse experience and/or profession (i.e., registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices (1) involve horses; (2) don't involve horse, and; (3) laypersons who do not fit into the other two groups). In this phase, the three groups of participants were shown the retained 10 videos (in randomized orders) and asked to describe the affective states of the horses in each. The terms generated by the three groups of participants were then compared (using a code developed here-within) to the answer keys created from the expert participant's terms. This resulted in a score for each individual participant of the three groups (and a mean score for each group) that was reflective of their assessment accuracy through which each prediction could be tested.

Methods

As just described, this study involved a two-phased approach. In the first (instrument development) phase, twelve experts were recruited to establish the validity of the EAAAI. In the second (prediction testing) phase, participants were recruited from three groups (EAMHPs, non-equine assisted mental health professionals [NEAMHPs], and laypeople).

ACCURACY OF HORSE AFFECT ASSESSMENTS

This study was approved by the Research Ethics Board of the University of Alberta (REB #: Pro00121752) for the ethical use of humans in research. The study was hosted online using the Survey Qualtrics platform and was open from February 2022 to July 2022 (the instrument development phase) and from November 2022 to July 2023 (the prediction testing phase). This study received funding from the Social Sciences and Humanities Research Council (SSHRC) in 2021 via a Joseph Armand Bombardier Canada Graduate Scholarship-Master's (CGS-M) award.

Instrument Development Phase: Equine Affect Assessment Accuracy Indicator

Operant Definitions

Accuracy is calculated by the EAAAI via comparing the terms and categories generated by the three groups of participants (in the prediction testing phase) against the correct and incorrect categories identified for each video, as determined by the answer keys created from the expert participant's answers in the instrument development phase. The terms below have been presented in order of relevance to one another, rather than alphabetically.

Accuracy

Used within this paper to refer to how similar terms/categories generated by the three group participants (in the prediction testing phase) were to those generated by the experts (in the instrument development phase). Accuracy should therefore be understood to mean "accuracy as compared to the expert group" in this study. Accuracy was calculated using the scoring method described in detail in Appendix C.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Terms

Refers to the descriptors of demeanor participants were instructed to generate to describe the horse they had just seen after watching each of the presented video clips. These terms could include words such as “happy,” “sad,” “nervous,” and so on.

Categories

Expert participant terms were grouped into categories based on their definitions and synonyms (the list of these categories and their associated terms can be found in Appendix D). A total of 35 categories were developed, 34 containing internally related descriptive terms of demeanour and one which contained terms not defined as descriptors of demeanour (e.g., “pregnant”). The process of categorization is explained in greater detail in Appendix E.

Correct Categories

The set of categories identified by the expert participants (via their terms) for each individual video were defined as the correct categories for that video. So, if a participant in the prediction testing phase viewing video X produced a term that fit into one of the correct categories identified by the experts in the instrument develop phase, for video X, the prediction testing phase participant would be awarded points (see Appendix C).

Incorrect Categories/Terms

Any categories (and their associated terms) outside of the set of correct categories identified by the experts for each individual video represent incorrect categories/terms for that specific video. So, if a prediction testing phase participant viewing video X produced a term that did not fit into one of the correct categories identified by the experts in the instrument development phase, for video X, the prediction testing phase participant’s

ACCURACY OF HORSE AFFECT ASSESSMENTS

term would belong to an incorrect category, and they would receive negative points (see Appendix C).

Participants: The Experts

Population of Interest and Sample

In the instrument development phase, English speaking adults (over the age of 18), of all genders, living in Canada and the USA who represent socially accepted experts of horse behaviour comprised the population of interest. The sample consisted of 12 participants who met this rigorous criteria: four large animal veterinarians operating out of an equine focused clinic, two professional horse trainers who use traditional/pressure–release methodologies in their training, two professional horse trainers who use positive reinforcement/clicker–training methods, and four equine behaviourists (two of which were certified with the American Veterinary Society of Animal Behaviour and two who were certified with the International Association of Animal Behaviour Consultants). The sample consisted of 11 individuals identifying as female and one individual who identified as male.

Exclusion and Inclusion Criteria

Participants under the age of 18 and/or who did not belong to the above-mentioned professions were ineligible to participate in the instrument development phase. Participants were included regardless of sex, gender, ethnicity, culture, and race as there is no research nor incentive to exclude participants based on any of these characteristics.

Sampling Procedure

Eligible potential expert participants (achieved sample size $n = 12$; intended sample size $n = 12$) were identified through internet searches related to the professions

ACCURACY OF HORSE AFFECT ASSESSMENTS

and countries (Canada and the USA) of interest. Personal and organizational websites were then carefully reviewed and if eligibility was confirmed those individuals/organizations were contacted via details listed publicly on the website.

All recruitment correspondents were approved by the ethics committee. Of the sample approached 13% completed participation in the study.

EAAAI Development Study Design

Independent and Dependent Variables

In this phase of the study, the terms participants used to describe the affective states of the observed horses were the independent variables of interest. In turn, inter-observer consensus was the dependent variable.

How Data Was Measured and Recorded

The study was distributed using online survey software (i.e., Qualtrics) to maximize eligible participant coverage while minimizing costs and time spent collecting data and eliminating the need for in-person testing due to risks associated with COVID-19. In this phase, 20 video clips were shown, in randomized orders, to all the expert participants.

Stimuli: The Videos

Twenty video clips of horses performing spontaneously occurring behaviours with an average runtime of 63 seconds (range of 54 seconds – 1 minutes 23 seconds) were prepared for this phase. No horse was ever coerced to obtain footage. The length of the videos was limited in order to reduce the participant time commitment needed to complete the study, thereby hopefully increasing participation. The video lengths in the study align with video lengths used by Minero et al., (2016, 2018). Each video clip

ACCURACY OF HORSE AFFECT ASSESSMENTS

presented footage of either one individual horse, or in instances where several are visible, the participant was explicitly told to focus on one specific, easily observed animal. Within each clip the horse being observed would demonstrate a collection of behaviours that were considered similar (for example, a video in which a horse goes from a sleeping state to startled/alert state would not be shown, as these behaviours are dissimilar). Video footage has the unique advantage over live observation of ensuring the observed animal's behaviour remains consistent across the participants.

Procedure

Before being shown the videos, participants were: presented with a consent form; asked to complete a demographic form (collecting information related to age, gender, and profession type; see Appendix F); asked to briefly describe their attitudes or philosophies towards horse training (this was done to ensure the professional horse trainer samples were correctly represented); and then were provided with instructions regarding completion of the study activity. The instructions were adapted from previous studies that have used the Qualitative Behavioural Assessment process (described in the Results: Qualitative Behavioural Assessment section; Clarke et al., 2016; Fleming et al., 2013; Gronqvist et al., 2017; Minero et al., 2009, 2016, 2018; Wemelsfelder et al., 2000, 2001) and can be seen in Appendix G.

Participants were then presented one video at a time, underneath which they were provided room to generate a minimum of one to a maximum of ten term names, term descriptions, and asked to rate the prevalence of the term they selected (described in more detail in the Results: Qualitative Behavioural Assessment section and in Appendix H).

ACCURACY OF HORSE AFFECT ASSESSMENTS

More in-depth details regarding the procedures used in this study can be found in Appendix I.

Prediction Testing Phase

Three predictions were made regarding the total assessment accuracy scores expected to be observed among the three groups of participants. First, it was predicted that the EAMHPs would demonstrate greater accuracy in assessing the horses than the NEAMHPs as the former were likely to have a greater degree of horse experience. Second, it was predicted that the EAMHP group would demonstrate greater accuracy in assessing the horses than laypeople with low levels of horse experience. Third, it was predicted that people with high levels of horse experience (regardless of profession) would be more accurate than those people with none/low levels of horse experience.

Participants: The Three Groups

The population of interest was English speaking adults (over the age of 18), of all genders, living in Canada and the USA. This study was specifically interested in the comparison of three groups: (1) registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices involve horses; (2) registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices don't involve horses; (3) and laypersons who do not fit into the other two groups. Participants with various levels of horse experience additionally comprised the population of interest.

Inclusion Criteria. Participants were included regardless of sex, gender, ethnicity, culture, and race as there is no research nor incentive to exclude participants based on any of these characteristics. Participants were also included regardless of their

ACCURACY OF HORSE AFFECT ASSESSMENTS

level of experience with horses, as the impact of horse experience on assessment accuracy was a factor of interest (Gronqvist et al., 2016, 2017; Krüger, 2007; McBride & Long, 2001; Merkies et al., 2018).

Exclusion Criteria. Anyone under the age of 18 was excluded, as it is not possible for such an individual to be a registered, licensed, and/or certified psychologist, counsellor, or mental health professional. While the layperson group does not require these qualifications, to keep the three groups comparable demographically, the minimum age limit was also extended to this group. Participants who identified themselves as unregistered, unlicensed, and/or uncertified psychologists, counsellors, or mental health professionals were excluded from the study.

Sampling Procedure. Laypersons (achieved sample size $n = 94$; intended sample size $n = 50$) were recruited through self-selection using the University of Lethbridge's SONA system and the posting of recruitment ads to public and private social media groups and accounts.

Both registered, licensed, and/or certified EAMHPs (achieved sample size $n = 55$; intended sample size $n = 50$) and NEAMHPs (achieved sample size $n = 51$; intended sample size $n = 50$) were also recruited through self-selection via the posting of recruitment ads to public and private Facebook groups and identified using internet searches, where upon eligible potential participants were then contacted and invited to take part in the study. Regarding the EAMHPs, of the sample approached ($n = 386$), 14% completed participation in the study.

Additionally, potential NEAMHP participants were also identified using PsychologyToday.com and through public provincial psychologist directories. Of the

ACCURACY OF HORSE AFFECT ASSESSMENTS

NEAMHP sample approached ($n = 745$), 6% completed participation in the study. Details regarding the exact procedure used to recruit participants can be found in Appendix J.

Prediction Testing Phase Design

Independent and Dependent Variables

In the prediction testing phase the independent variables were profession (three levels, categorical) and horse experience (comprised of self-rated level of horse experience [a four-level ordinal variable] and horse experience scores [discrete]). The dependent variable was the total assessment accuracy score participants earned for the 10 videos they watched as determined by the EAAAI.

How Data Was Measured and Recorded

The study was once again distributed using online survey software (i.e., Qualtrics). Participants were again presented one video at a time, underneath which they were asked to generate one to ten: term names, term descriptions, and ratings of prevalence for the terms they selected (described in more detail in the Results: Qualitative Behavioural Assessment section and in Appendix H). More in-depth details regarding the procedures used in this study can be found in Appendix I.

Procedure

Once again, participants were presented with a consent form and were additionally asked to confirm whether or not they lived in Canada or the USA. If the participant consented and indicated that they lived in Canada or the USA, they were presented with a demographics form (see Appendix K) which gathered information regarding their age and gender. Participants were asked further questions regarding their profession type (the answers to which were used to assign them to the appropriate group;

ACCURACY OF HORSE AFFECT ASSESSMENTS

details about this process can be found in Appendix I), were presented with an additional horse experience questionnaire, and were asked to self-rate their level of horse experience. Horse experience was scored using the creation of a code to allot points based on the answers selected by the participants to the horse experience questionnaire (see more detailed below and in Appendix L).

Instructions

The delivery of the instructions was improved for this phase through the implementation of an additional instructional video (runtime of 3 minutes and 29 seconds) accompanied by written instructions. This video introduced the participant to the layout of the study activity and reviewed the same information as included in the written instructions presented in the instrument development phase (see Appendix G). A timer was also implemented so that the participant could not advance until the duration of the instructional video had lapsed. After the instructions were presented, participants were shown (one at a time and in a randomized order for each participant) 10 videos.

Changes to the Prediction Testing Phase's Activity

In this phase, two changes were made to improve the activity's user-friendliness: (1) The videos were pinned to the top of their pages so that as the participant scrolled down to fill in the term names and term descriptions the video would always remain visible. This allowed participants to easily re-watch, pause, and play the videos as they completed the activity; and (2) timers were implemented on every page of the study activity which contained a video so that the "Next" button only appeared after the duration of the associated video had lapsed. This was to ensure participants had fully read and/or listened to the instructions before advancing to the study activity. In-depth details

ACCURACY OF HORSE AFFECT ASSESSMENTS

of the procedure associated with this, and the instrument development phase can be found in Appendix I.

Results

Refining the Experts Survey into the Equine Affect Assessment Accuracy Indicator

Following the collection of data, the results of the instrument development phase were prepared for a Generalized Procrustes Analysis to be run. The purpose for running the Generalized Procrustes Analysis was to produce the associated Residuals by Object table and graph, in order to identify the top 10 (of the 20) videos presented for which the expert participants had the highest degree of consensus. The associated terms produced by the experts could then be retained for use as answer keys, from which a scoring code could be created – thus allowing for the development of the EAAAI.

Preparing the Data

To prepare the data, terms produced by the experts were sorted into 34 categories, and a code was created which either automatically re-labelled the terms with the associated category name or flagged more ambiguous terms so they could be hand-reviewed and assigned appropriately. The step-by-step details involved in the process of categorizing terms can be found in Appendix E.

As one of the stated intentions of this phase was to acquire answer keys for the retained 10 videos, a categorical approach (over a term-based approach) for awarding points was used. This decision was made to avoid penalizing participants from the three groups in the prediction testing phase who were anticipated to generate terms that were different—yet correct—from those produced by the expert participants. The importance of categories becomes evident when we consider that it would be inappropriate to

ACCURACY OF HORSE AFFECT ASSESSMENTS

penalize a participant from the prediction testing phase for producing, for example, the term “joyful” to describe a horse in video X, when the expert participants only produced the term “delighted.” In this instance, the prediction testing phase participant would have correctly identified the horse’s affective state in accordance with the expert participant, and yet they would be penalized as their term did not precisely match the experts. However, if “joyful” and “delighted” instead belonged to the same category, then the prediction testing phase participant could be awarded the appropriate points for identifying the correct category and the similarity of their answer to that of the experts could be honoured.

Thus, terms produced by the expert participants were categorized to permit this method of awarding points to be used in the prediction testing phase of the study. A complete list of the categories and the terms associated with each can be found in Appendix D.

Qualitative Behavioural Assessment

The Qualitative Behavioural Assessment (QBA) process was utilized in the development of the EAAAI in this study. A selection of research has shifted towards a “whole animal” approach when assessing animal welfare through the use of the QBA (Wemelsfelder et al., 2000). In a *Qualitative Behavioural Assessment*, a person looks at an example of animal behaviour (either via watching a live animal or footage of an animal; Fleming et al., 2013; Wemelsfelder et al., 2001) and either generates their own terms to describe the affective state(s) of the animal viewed (this is known as *free-choice profiling*) or selects the terms from a provided list of descriptive words (known as the *fixed-list method*; Minero et al., 2009; Wemelsfelder et al., 2000, 2001). The goal of the

ACCURACY OF HORSE AFFECT ASSESSMENTS

QBA is to have the client describe the overall demeanor or emotional state of the animal they are observing, using descriptive terms. In this study, free choice profiling was used (for more details about the differences between these two methods see below and Appendix M).

Fixed Lists. A fixed list presents every participant in the study with the same set of species-specific descriptive terms to choose from in order to describe the animal presented in the stimuli (Clark et al., 2016). These lists are usually developed through a focus group of experts, who generate new terms, debate ones from the literature, and eventually reach consensus on the fixed list to be used (Clark et al., 2016; Gronqvist et al., 2017; Minero et al., 2016, 2018). Sometimes these lists may also be informed by an existing *ethogram* (Hall & Heleski, 2017; McDonnell & Haviland, 1995), which is “a formal list of a species’ behavioral repertoire...[that] may be a complete list of all behaviors or it may focus on particular functional classes of behaviors” (Grier, 1984, p. 56 & 69). For an example of a demeanor-focused horse ethogram please see table 6 in Torcivia and McDonnell (2021).

Free Choice Profiling. Free choice profiling (FCP), on the other hand, involves having participants generate their own terms to summarize the animal’s demeanour using expressive descriptors (e.g., “angry;” Clark et al., 2016). The terms are then defined by the participant and scored using a Visual Analogue Scale (described below). FCP has been used in many QBA studies across many species, including pigs (Clarke et al., 2016; Rutherford et al., 2012; Wemelsfelder et al., 2000, 2001), dogs (Arena et al., 2017; Walker et al., 2009), cattle (Rousing et al., 2006; Stockman et al., 2012, 2013; Vindevoghel et al., 2019), sheep (Fleming et al., 2015; Masłowska et al., 2020; Wickham

ACCURACY OF HORSE AFFECT ASSESSMENTS

et al., 2015) and of course, horses (Fleming et al., 2013; Hintze et al., 2017; Minero et al., 2009, 2018; Napolitano et al., 2008).

Decision to use FCP. FCP is arguably more reflective of real-world therapy scenarios, where an equine assisted therapist must make decisions in the moment to maintain and protect the welfare of the horse during a session based on their experience and knowledge. It would be disruptive as well as unrealistic to expect a therapist to pause a session and pull out a fixed list to assess the horse with. Further, this study is interested in the abilities of EAMHPs to make accurate, real-time assessments of their co-therapists so that they can respond appropriately in the moment. For this and the reasons presented in Appendix M, the FCP method was selected.

Visual Analogue Scale. Regardless of the method of term selection employed for a QBA, the participant is then asked to rate the prevalence of the term(s) they have generated/selected on a Visual Analogue Scale. A *Visual Analogue Scale* (VAS) is a scale comprised of two poles labeled ‘minimum’ and ‘maximum’ with a continuous uncategorized line between them (Minero et al., 2009). Using paper delivery the VAS score is “recorded as the measure of the distance in millimeters between the left ‘minimum’ point of the scale and the point where the observer’s [tick] crosses the line” (Minero et al., 2009, p. 6). In the present study the VAS took the form of a digital slider scale ranging from 0 to 100 and the score was derived from the number at which the participant dragged the slider to. The slider was always presented in the default position of 50 but was not recorded as such until the participant interacted with it. For the purposes of this study, the left pole of these sliding scales was labelled “minimum (this expressive quality is absent)” and the right pole was labelled “maximum (this quality

could not be present more strongly)” (Minero et al., 2016, p. 149). See Appendix H for more details regarding the use of a VAS.

Generalized Procrustes Analysis. After the data was gathered, the multivariate statistical technique of the Generalized Procrustes Analysis was conducted using the terms produced by the participants for each video and their associated VAS scores (Fleming et al., 2013; Mellor, 2014; Wemelsfelder et al., 2000, 2001). As Wemelsfelder et al. (2001) described, the

[Generalized Procrustes Analysis] can be thought of as a pattern detection mechanism and is based on the assumption that even if observers use different variables (terms) for measurement, the distances between samples [horses] as specified by the various measurements are comparable, *because the samples are the same*. In other words, [Generalized Procrustes Analysis] takes for granted that measurement patterns all dealing with the same samples will converge and is designed to compute the coordinates of the convergent configuration (the so-called ‘consensus profile’). Thus, [Generalized Procrustes Analysis] detects the level of consensus between observer assessment patterns not on the basis of fixed variables (terms), but on the basis of the (multi-dimensional) inter-sample distances specified by each observer (i.e., how each observer uses his/her terms to score [horses]). (p. 198)

The level of consensus produced by the Generalized Procrustes Analysis (presented in a Residuals by Object table and graph) was the result of interest in this phase. By observing the residuals it could be determined for which 10 of the 20 videos the experts most agreed on the affective states of the horses presented. In turn those 10 videos were then used in the prediction testing phase.

The Use of The QBA in Research

The QBA has been used by researchers to investigate such topics as the health and welfare of endurance horses competing in a 160–km race (Fleming et al., 2013), the horse-human relationship in connection to the animal’s emotional state (Minero et al.,

ACCURACY OF HORSE AFFECT ASSESSMENTS

2018), and to examine the effect intensive early handling has on the responses of yearlings toward unfamiliar humans (Minero et al., 2009), to name a few.

The QBA's Purpose. The QBA process takes into account that the affect or mental state of an animal cannot be determined based on isolated behaviours. Instead, the whole animal must be taken into consideration, with the focus less on what it is doing and more on how it is doing it (Fleming et al., 2013; Mellor, 2014; Wemelsfelder et al., 2000, 2001). The QBA process involves the observer (1) considering the entire animal (its posture, behaviours, and movements), (2) summarizing these observations using expressive descriptions of demeanour, for example “happy” or “upset,” and (3) rating the prevalence of the descriptors using a VAS (Fleming et al., 2013; Mellor, 2014; Wemelsfelder et al., 2000, 2001). The QBA operates on the belief that the subtle details discerned when taking the whole animal into consideration may be missed when a person only assesses an animal based solely on isolated behaviours, and thus increasing the likelihood that errors in detecting shifts in demeanour and welfare may occur (Fleming et al., 2013).

Validity and Application. To date, the majority of studies which have incorporated the QBA process have focused on its validity and application (see Clarke et al., 2016; Fleming et al., 2013; Gronqvist et al., 2017; Minero et al., 2009, 2016, 2018; Wemelsfelder et al., 2000, 2001). The QBA has been used in many studies in the assessment of many species, including horses, and has been found to hold merit as a welfare assessment tool (Fleming et al., 2013; Mellor, 2014; Wemelsfelder et al., 2000, 2001). The scores generated from the QBA can be correlated with both “quantitative

ACCURACY OF HORSE AFFECT ASSESSMENTS

measures of behaviour” and “with physiological indicators relevant to welfare” (Fleming et al., 2013, p. 2).

The QBA is a robust, flexible, and easily implemented method for assessing animal welfare both in the field and after-the-fact (Fleming et al., 2013; Wemelsfelder et al., 2001). For example, Fleming et al. (2013) used the QBA to investigate whether equine health and welfare was compromised before, during, and after the horses had participated in a strenuous 150 km endurance race using video footage. While there is some criticism among the behavioural studies that such analyses cannot be done in the absence of background knowledge of a horse’s temperament and its impacts on the animal’s interactions with its environment, the findings of Fleming et al. (2013) demonstrated that the QBA is a valuable additional tool for assessing and researching equine temperaments without the need for background temperament knowledge. This finding is important for the current study as neither the expert participants (instrument development phase) nor the three groups of participants (prediction testing phase) were given any contextual or temperament information regarding the horses presented in the videos.

Generalized Procrustes Analysis: Residuals by Object

The Residuals by Object table and graph produced by the Generalized Procrustes Analysis identified the number of residuals each video possessed (see Table 1). Residuals represent how far away the object (i.e., video) was from its associated consensus matrix whereby the objects (i.e., videos) with the smallest number of residuals are taken to represent those videos for which there is a greater consensus among the experts (Dijksterhuis & Gower, 1991; Gower, 1975; Mauricio et al., 2016).

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 1

Instrument Development Phase: Results of the Generalized Procrustes Analysis: Residuals by Object

Object	Residuals	Object	Residuals
Video A	36753.312	Video K	26775.390
Video B	38724.366	Video L	23183.969
Video C	47385.203	Video M	34319.341
Video D	21344.337	Video N	46596.500
Video E	38066.964	Video O	16833.301
Video F	26474.178	Video P	28332.415
Video G	28303.945	Video Q	30180.060
Video H	22086.361	Video R	29669.900
Video I	26069.511	Video S	39525.280
Video J	36453.879	Video T	40109.808

Note. Typically, those objects with the smallest number of residuals have the greater degree of consensus among the observers. The range of residuals presented here are reflective of variation among the experts regarding the number of terms each one produced per video. The fact that they begin in the thousands therefore does not invalidate the findings.

There exist no guidelines for how big is too big (or how small is too small) when it comes to residuals. The residuals calculated by the Generalized Procrustes Analysis here range in the thousands as the same number of terms were not produced by every participant for every video (as participants were instructed to produce between 1 and 10 terms per video). These observed residuals are therefore reflective of the unequal number of terms generated between participants (resulting in missing values) and are not indicative of low consensus.

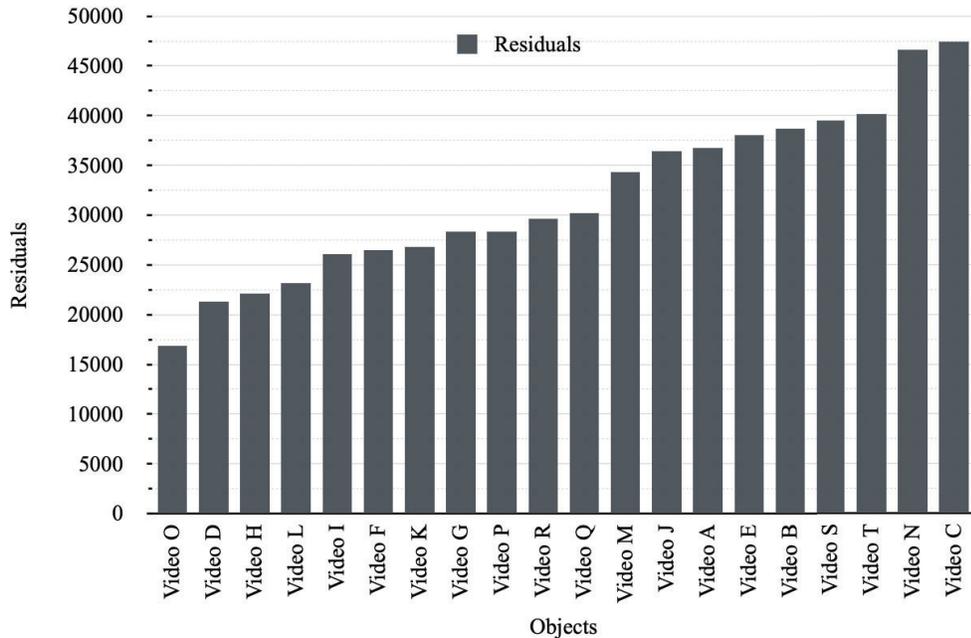
The results of the generalized Procrustes Analysis' Residuals by Objects table indicates that videos O, D, H, L, I, F, K, G, P, and R had the greatest level of expert

ACCURACY OF HORSE AFFECT ASSESSMENTS

participant consensus (in decreasing order), and videos C, N, T, S, B, E, A, J, M, and Q had the lowest degrees of consensus (see Figure 1).

Figure 1

Instrument Development Phase: Generalized Procrustes Analysis: Videos Organized by Residuals



Note. This is a graph of the residuals by object (videos) found in the instrument development phase using a Generalized Procrustes Analysis with the videos presented in ascending order.

Video Selection Informed by the Generalized Procrustes Analysis

Additionally, the categories and VAS scores produced by the participants were hand-reviewed as well. This raised questions about the Generalized Procrustes Analysis' identification of videos J and S as having low consensus and videos D and G as having high consensus. After carefully reviewing the answers produced by the experts for these four videos the decision was made to exclude videos D and G and replace them with videos J and S. The Generalized Procrustes Analysis seemed to have mistaken the use of multiple (but applicable) categories by the experts to describe the horses in videos J and S as a reflection of low consensus, when in fact this was not the case. For a thorough

ACCURACY OF HORSE AFFECT ASSESSMENTS

discussion regarding the decision to exclude videos D and G and include videos J and S please see Appendix N.

The final 10 videos selected for presentation in the prediction testing phase were as follows: videos F, H, I, J, K, L, O, P, R, and S. As they were determined to have the greatest interobserver consensus these 10 videos were retained, along with the terms produced by the expert participants (which were used to create the associated answer keys) for each. Using these in conjunction with the code created for scoring described in the next section, the EAAAI was developed.

Code Development

The terms generated by the expert participants in the instrument development phase were categorized to construct answer keys for each of the 10 retained videos (and any new terms generated in the prediction testing phase were also assigned to these existing categories). Based on the contents of these categories a code was developed which either automatically re-labelled the terms with the associated category name or flagged ambiguous terms so they could be hand-reviewed and assigned appropriately.

A second code was also developed which assigned weighted values to correct categories and a fixed negative value to incorrect terms identified by the three groups of participants in the prediction testing phase. The weighting of correct categories was informed by the answer keys created from the terms the expert participants generated for each video. For a detailed description of how the EAAAI was used to score the accuracy of participants from the three groups in the prediction testing phase please refer to Appendix C.

ACCURACY OF HORSE AFFECT ASSESSMENTS

The Equine Affect Assessment Accuracy Indicator

The outcome of the instrument development phase was the creation of the Equine Affect Assessment Accuracy Indicator. The purpose of the EAAAI is to permit the scoring of a participant's accuracy when assessing the affective states of horses based on video footage.

To use the EAAAI, first the horse experience questionnaire is administered to the participant. Next, 10 video recordings of horse behaviour are presented one at a time, in a randomized order, to a participant on a computer. The participant is asked to generate 1–10 terms to describe the affective state of the horse they are observing. The participants are also asked to briefly define the terms they generate, and to rate the prevalence of each term (i.e., how [happy] was the horse on a scale of 0 to 100?) using a digital VAS that ranges from 0 to 100, with the default position of the marker set to 50 (but not counted as such unless the participant interacts with the marker).

Lastly, two codes are used, the first of which either automatically re-labels participant terms in the raw data to the category each term is associated with, or flags terms which require further review. Once all terms have been categorized, the second code then grades each participant. A participant can obtain a total assessment accuracy score as high as +50.0 or as low as -50.0.

Prediction Testing Phase

Planned Statistical Analysis

The objectives of the prediction testing phase were to investigate whether (a) a person's degree of horse experience and/or (b) their profession (EAMHPs, NEAMHPs, or

ACCURACY OF HORSE AFFECT ASSESSMENTS

laypeople) has an impact on their ability to accurately assess horse affect. It was predicted that:

1. The EAMHPs would be more accurate than the NEAMHPs.
2. The EAMHPs would be more accurate than laypeople with low levels of horse experience.
3. People with high levels of horse experience (regardless of profession) would be more accurate than people with none/low levels of horse experience.

To this end four main analyses were conducted. First, a correlation was run to assess the relationship between participant's self-rated level of horse experience and the horse experience scores they earned for their answers to the horse experience questionnaire. Second, a between subjects one-way Welch's analysis of variance (ANOVA) was used to assess any main effect of profession. Third, a between subjects one-way ANOVA was conducted to assess any main effect of self-rated levels of horse experience, and forth, a multiple regression was run to assess the variation in the total assessment accuracy scores that could be explained by the assigned horse experience scores and profession.

Preparing the Data. The answers the participants provided on the horse experience questionnaire were then scored in accordance with the process outlined in Appendix L. These scores could range from 0–71 (where the higher the score, the more horse experience a person was determined to possess). The terms the participants generated for each video were then processed using the EAAAI (see appendix C). A detailed review of the procedure used to prepare the results of the horse experience questionnaire and EAAAI for statistical analysis can be found in Appendix O.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Composition of the Three Group Samples

A review of the mean assessment accuracy scores across the three groups revealed that laypeople had the highest total assessment accuracy scores, followed by the equine assisted mental health professionals. The NEAMHPs had the lowest scores.

Table 2

Descriptive Analysis of the Prediction Testing Phase Participant Sample

Demographic characteristic	Layperson		EAMHP ^a		NEAMHP ^b	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Age						
18-24	32	34	1	1.8	-	-
25-34	14	14.9	13	23.2	14	27.5
35-44	15	16	19	33.9	15	29.4
45-54	10	10.64	11	19.6	8	15.7
55-64	18	19.15	9	17.9	9	17.6
65-74	5	5.31	2	3.6	5	9.8
75+ ^c	-	-	-	-	-	-
Gender						
Female	82	87.23	51	92.9	38	74.5
Male	10	10.64	3	5.4	11	21.6
Non-binary / Other	2	2.12	-	-	2	3.9
Prefer not to Say	-	-	1	1.8	-	-
Self-rated Level of Horse Experience						
None	17	18.1	-	-	20	39.2
Low	23	24.5	2	5.4	25	49
Medium	20	21.3	18	32.1	4	7.8
High	34	36.2	35	62.5	2	3.9

Note. *N* = 200 participants (consisting of 94 laypersons, 55 EAMHPs, and 51 NEAMHPs).

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

^c Zero participants reported being 75 or older

In total, 200 participants were included in the statistical analysis for the prediction testing phase. The descriptive characteristics of these participants are summarized and presented in Table 2. The majority of participants identified as female. While most participants in the NEAMHP group identified as having none to a low level of horse

ACCURACY OF HORSE AFFECT ASSESSMENTS

experience (88%), the laypeople group represented a more even distribution of those with no–low experience (43%) to those with a medium–high level of experience (57%).

By its nature it was anticipated that none of the EAMHP participants would report having no experience with horses, and this was true to the observed results. However, it was surprising to find that two of the EAMHPs reported having a low level of horse experience (5.4%). Upon examination of the answers provided by these two participants, it was found that both reported using a therapeutic non-riding/ground model and an equine-facilitated psychotherapy model (with one participant additionally saying they use horses to help facilitate relationship counselling) in their practice.

The exclusive use of an equine-facilitated psychotherapy model by these two participants could explain why they identified themselves as having a low level of experience, as this model is most often comprised of four parties: the horse, the client, and one expert who handles the animal and is a credentialed equine professional, and another expert who is the therapist (Bona, 2020; Nilson, 2004). If this was the case, these two participants may have felt they lacked horse experience because they may not have been directly handling the animal during therapy sessions.

Correlation of Self-rated Horse Experience and Horse Experience Scores

To quantitatively assess a participant's level of horse experience, they were asked to fill out a horse experience questionnaire before starting the study activity (see Appendix L). The results of this questionnaire and the participants' self-rated levels of horse experience were analysed with a Kendall's tau-b correlation.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Findings

There was a statistically significant, strong, positive association observed between the level of horse experience participants perceive they have and the scores they were assigned based on their answers to the horse experience questionnaire, where an increase in one was associated with an increase in the other. Further, even after splitting the data by profession and performing the analyses again, this statistically significant, strong,

Table 3

Correlations Related to Horse Experience Scores and Self-rated Levels of Horse Experience

Variable	1	2
Overall ^a		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.742**	—
Layperson ^b		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.703**	—
EAMHP ^c		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.533**	—
NEAMHP ^d		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.647**	—

^a A Kendall's tau-b correlation determined that the relationship between the prediction testing phase participants' ($n = 200$) self-rated levels of horse experience and the horse experience scores they received had a statistically significant, strong, positive association ($\tau_b = .742, p < .001$).

^b After splitting the data by profession, a second Kendall's tau-b correlation determined that the relationships between the prediction testing phase participants' self-rated levels of horse experience and the horse experience scores were statistically significant, strong, and positively associated for the laypeople group ($\tau_b = .703, p < .001$), the EAMHP group ($\tau_b = .533, p < .001$), and the NEAMHP group ($\tau_b = .647, p < .001$).

^c Equine assisted mental health professionals

^d Non-equine assisted mental health professions

* $p < .05$. ** $p < .01$

ACCURACY OF HORSE AFFECT ASSESSMENTS

positive relationship was still observed in all three groups. The results of these two correlations can be seen in Table 3.

Between Subjects One-Way Welch's ANOVA: Main Effect of Profession

One outlier³ was detected via inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. No extreme outliers were observed. The normal distribution of the total assessment accuracy scores for all three groups was confirmed through the Shapiro-Wilk's test ($p > .05$), and visual inspection of their histograms and Normal Q-Q Plots. Please see Appendix P for the observed skewness, kurtosis, and associated standards of error. The homogeneity of variance between the three groups of participants was violated, as reported by a Levene's test based on the median, $F(2, 197) = 3.674, p = .027$.

Findings

A between subjects one-way Welch's ANOVA was therefore conducted to analyse any main effect of profession type on the dependent variable of total assessment accuracy scores. There was a statistically significant main effect of profession where Welch's $F(2, 118.767) = 10.682, p < .001, \eta^2 = .074$, with an observed power of .951 ($\alpha = .05$). The means and standard deviations of the three groups of participants in the prediction testing phase are reported in Table 4.

³ The outlier identified belonged to an EAMHP whose total assessment accuracy score was = +14.6368. To account for this outlier, the participant's total assessment accuracy score was changed to the next largest value plus 0.01 to maintain its rank as the highest value in the data set (changed to $13.295 + 0.01 = 13.305$). Once this change was made no other outliers were detected.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 4

Average Total Assessment Accuracy Scores Across All Professional Groups

Profession	N	Mean	Standard Deviation
Layperson	94	5.77	3.51
EAMHP ^a	55	5.12	3.53
NEAMHP ^b	51	3.52	2.43

Note. Presented here are the means and standard deviations for the average total assessment accuracy scores for each of the three profession groups analyzed in the prediction testing phase. A statistically significant main effect of profession was found *Welch's* $F(2, 118.767) = 10.682, p < .001, \eta^2 = .074$ with an observed power of .951 (alpha = .05).

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

A post-hoc Games-Howell found that both the laypeople and EAMHPs scored significantly higher than the NEAMHPs. The analysis indicated that the NEAMHP group had significantly lower total assessment accuracy scores compared to both the equine assisted mental health professional group ($p = .019$), with a mean decrease of -1.59 points, 95% CI [-2.97, 0.22], and the laypersons group ($p < .001$), with a mean decrease of -2.25 points, 95% CI [-3.42, -1.07]. With a mean increase of 0.65 points, there was no statistically significant difference ($p = .515$) in total assessment accuracy scores between the laypeople and EAMHP groups 95% CI [-0.75, 2.06].

It is interesting that no significant differences were observed between the laypeople and the EAMHPs. As the proportion of participants who self-rated themselves as having a medium or high level of horse experience was much greater for the EAMHPs (94.6%) than the laypeople (57.5%), a difference in scores was anticipated in this analysis.

One-Way ANOVA on Self-Rated Horse Experience Levels

A between-subjects one-way ANOVA was conducted to analyse any main effect of self-rated levels of horse experience on the dependent variable of total assessment accuracy scores. Five outliers were detected in the data as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. No extreme outliers were observed. The outliers identified belonged to one participant who rated themselves as having no horse experience⁴ and four participants who rated themselves as having a low level of horse experience.⁵ Once these changes were made no other outliers were detected.

The normal distribution of the total assessment accuracy scores for all four groups was confirmed through the Shapiro-Wilk's test ($p > .05$), and visual inspection of their histograms and Normal Q-Q Plots. See Appendix Q for a table of the observed skewness, kurtosis, and associated standards of error. There was homogeneity of variance, as assessed by a Levene's test, based on the median, $F(3, 197) = 0.998, p = .395$.

⁴ With a total assessment accuracy score of = +13.0893 points. To account for this outlier their score was changed to $11.1628 + 0.01 = 11.1728$ points, to maintain its rank as the highest value in the no experience group.

⁵ Whose total assessment accuracy scores were changed from: +13.2953 to +9.2306; +12.3896 to +9.2406; +12.0399 to +9.2506; and +11.1507 to +9.2606 points, to maintain their ranks and order as the highest values in the low level horse experience group.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 5

Average Total Assessment Accuracy Scores Across All Self-Rated Horse Experience Levels

Self-Rated Level of Horse Experience	N	Mean	Standard Deviation
None	37	3.92	3.34
Low	50	4.41	2.63
Medium	42	4.93	3.45
High	71	5.9	3.32

Note. Presented here are the means and standard deviations for the average total assessment accuracy scores for each of the four self-rated levels of horse experience analyzed. A statistically significant main effect of self-rated horse experience was found $F(3, 196) = 3.832, p = .011, \eta^2 = .055$ with an observed power of .814 ($\alpha = 0.5$).

A post-hoc Games-Howell analysis indicated that both the participants who said they have no horse experience ($p = .024$ with a mean decrease of -1.97, 95% CI [-3.75, -0.20]) and those that said they have a low level of horse experience ($p = .035$ with a mean decrease of -1.49, 95% CI [-2.90, -0.07]) scored significantly lower than those participants who said they have a high level of horse experience.

Findings

There was a statistically significant main effect of self-rated level of horse experience $F(3, 196) = 3.832, p = .011, \eta^2 = .055$ with an observed power of .814 ($\alpha = .05$). A post-hoc Games-Howell analysis found that participants with none or a low level of horse experience were significantly less accurate than those with a high level of self-rated horse experience.

Specifically, the analysis indicated that both the participants who said they have no horse experience ($p = .024$ with a mean decrease of -1.97 points, 95% CI [-3.75, -0.20]) and those that said they have a low level of horse experience ($p = .035$ with a mean decrease of -1.49 points, 95% CI [-2.90, -0.07]) scored significantly lower than those participants who said they have a high level of horse experience. The means and standard deviations of the four self-rated horse experience levels are reported in Table 5.

There were no statistically significant differences between the scores of participants with no experience and (a) participants with a low level of experience ($p =$

ACCURACY OF HORSE AFFECT ASSESSMENTS

.884 with a mean increase of 0.49 points, 95% CI [-1.26, 2.24]), or (b) those with a medium level of experience ($p = .557$ with a mean increase of 1.01 points, 95% CI [-1.00, 3.01]). Secondly, there was no statistically significant difference between the scores of participants from the low level group and those from the medium level group ($p = .854$ with a mean increase of 0.52 points, 95% CI [-1.19, 2.22]). Lastly, there was no statistically significant difference between the scores of participants from the medium level group and those from the high level group ($p = .464$ with a mean increase of 0.97 points, 95% CI [-0.77, 2.70]).

In summary, significant differences in accuracy were only observed between (a) the no horse experience and the high horse experience groups and (b) the low horse experience and the high horse experience groups. These findings are not surprising, as differences in accuracy scores would be expected between those individuals with a great deal of horse experience compared to those with little to none.

Horse Experience Scores versus Self-rated Levels of Horse Experience

To accurately evaluate the effects profession and horse experience have on a person's ability to assess the affective states of a horse similarly to the experts it is important to determine which variable, the horse experience scores or the self-rated horse experience levels, is most representative of participant horse experience. Having examined the patterns of horse experience scores and the self-rated levels of horse experience, the former was determined to be a better indicator of participant horse experience and was therefore used in the following multiple regression analysis. For further details regarding this decision please see Appendix R.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Multiple Regression Analysis

A multiple regression was conducted to determine how much of the variation in the total assessment accuracy scores is explained by the horse experience scores and professional designations. The profession variable was dummy-coded for this analysis.

One outlier (see Footnote 3) was detected in the data as assessed by inspection of the case wise diagnostic table. No extreme outliers were observed. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.89. A scatterplot of standardized residuals versus standardized predicted values was also found to be linear. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. The assumption of collinearity was not violated. Lastly, the residuals were normally distributed as assessed by visual inspection of a normal probability plot.

Findings

Profession and horse experience scores, as a set, statistically significantly accounted for approximately 9.4% of the variation in the total assessment accuracy scores, $F(3, 196) = 6.76, p < .001$. The R^2 for the overall model was 9.4% with an adjusted R^2 of 8%, a small size effect according to Cohen (1988). The results of this multiple regression analysis are reported in Table 6.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 6

Results of Multiple Regression Analysis Pertaining to Horse Experience Scores and Profession

Variable	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	95% <i>CI</i>
Scored Horse Experience	0.028	0.014	2.063	0.04	[0.001, 0.056]
Reference Group: EAMHP ^a					
Constant	3.852	0.753	5.118	<. 001	[2.367, 5.336]
Laypeople	1.001	0.574	1.743	.083	[-2.134, 0.132]
NEAMHP ^b	-0.622	0.785	-0.793	0.429	[-0.926, 2.171]
Reference Group: Laypeople					
Constant	4.853	0.555	8.738	< .001	[3.757, 5.948]
NEAMHP	-1.623	0.638	-2.545	.012	[-2.881, -0.366]

Note. *N* = 200 participants (consisting of 94 laypersons, 55 EAMHPs, and 51 NEAMHPs). Profession was dummy-coded for this analysis. Profession and horse experience scores, as a set, statistically significantly account for variation in the total assessment accuracy scores, $F(3, 196) = 6.76, p < .001$. The R^2 for the overall model was 9.4% with an adjusted R^2 of 8%. Repetitious comparisons have been omitted.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

Horse experience scores can be seen to significantly ($p = 0.04$) positively predict an increase (+0.028 points) in accuracy scores, suggesting that the more experience a person has with horses the more likely they are to accurately assess their affective states. Regarding the professions, when the reference category is NEAMHP we see that a participant from this group with a horse experience score of 0 points would score 1.623 points (on total assessment accuracy) less than a layperson, which is a statistically significant difference ($p = .01$). When the reference category is changed to EAMHPs: (a) EAMHPs with a horse experience score of 0 points would have a conditional mean total assessment accuracy score of 3.852 points, (b) laypeople with a horse experience score of 0 points would score +1.001 points (on total assessment accuracy) more than the EAMHPs (a non-significant difference), and lastly (c) NEAMHPs with a horse

ACCURACY OF HORSE AFFECT ASSESSMENTS

experience score of 0 points would score -0.622 points (on total assessment accuracy) less than the EAMHPs (a non-significant difference). Therefore, these findings tell us that NEAMHPs score significantly lower than laypeople when horse experience is controlled for but no other significant differences between the groups exist.

Accuracy of Equine Affect Assessments

As previously defined, accuracy (i.e., assessment accuracy, assessment accuracy scores, equine affective assessment scores, total assessment accuracy scores, etc.) in this study refers to how similar terms generated by the three groups of participants in the prediction testing phase are to those generated by the expert participants in the instrument development phase. That is, accuracy can be understood to mean “accuracy as compared to the experts.”

Total Assessment Accuracy Scores

Total assessment accuracy scores have the ability to range from +50.0 points across all videos (+5.0 points per video) to -50.0 points across all videos (-5.0 points per video) where the higher the score, the more accurate the participant was. While these numbers help describe the psychometrics of the EAAAI, meaningful information is drawn from observing the actual scores.

How Well do the Experts Perform?

The answer keys, which the three groups of participants in the prediction testing phase were graded with, were created by retaining all of the terms (and their associated categories) generated by the 12 expert participants from the instrument development phase for each video. Therefore, situations occurred where, for example, one expert

ACCURACY OF HORSE AFFECT ASSESSMENTS

participant may have produced only one term (and thus distinguished only one category) out of all the correct categories identified across all of the experts for that video.

Table 7

Average Total Assessment Accuracy Scores Across All Professional Groups including Experts

Profession	<i>N</i>	Median	<i>M</i>	<i>SEM</i>	<i>SD</i>	Minimum	Maximum
Layperson	94	5.63	5.77	0.36	3.51	-1.34	13.3
EAMHP ^a	55	4.95	5.12	0.47	3.47	-1.71	13.31
NEAMHP ^b	51	3.37	3.52	0.34	2.43	-1.19	8.7
Experts	12	17.02	16.59	0.73	2.52	12.03 ^c	19.98

Note. Presented here are the medians, means, standard error of means, standard deviations, and minimum and maximum scores for the total assessment accuracy scores for each of the three groups of participants from the prediction testing phase and the expert participants from the instrument development phase.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

^c Because all of the terms generated by the expert participants in the instrument development phase of the study were used to construct the answer keys, against which the experts are being scored here in this exploratory analysis, this minimum score is not informative, as no expert was penalized for any of the terms they generated (and therefore they could never score negatively) unlike the three groups of participants from the prediction testing phase.

For exploratory purposes, the expert participants ($n = 12$) were scored using the EAAAI. Of the total assessment accuracy scores obtained two outliers were identified, one from the EAMHP group (see Footnote 3) and one from the expert participant group⁶

This analysis revealed that the mean total assessment accuracy score earned by the experts was 16.59 points, ($SD = 2.52$), with a maximum of 19.98 points and a minimum score of 7.409 points (which was the identified outlier [see Footnote 6], with the next

⁶ Who had a total assessment accuracy score of +7.409 which was changed to the next lowest value among the experts minus 0.01 to maintain its rank as the lowest value in the data set (changed to $12.042 - 0.01 = 12.032$).

ACCURACY OF HORSE AFFECT ASSESSMENTS

lowest expert score being 12.03 points). The means, standard deviations, and minimum and maximum scores for all four groups can be seen in Table 7.

The most valuable information presented here is that of the expert participants' mean and maximum total assessment accuracy scores. It is noteworthy that no expert obtained the maximum total assessment accuracy score of +50.0 points when they were scored using the EAAAI. The purpose of this exploratory analysis was to provide context for the total assessment accuracy scores observed across the three prediction testing phase groups. No other conclusions can or should be drawn from these presented findings.

Discussion

The results of the prediction testing phase of this study reveal interesting findings about the effects profession and experience have on a person's ability to assess the affective states of horses similarly to experts (affect assessment accuracy). It was predicted that (1) EAMHPs would be more accurate than NEAMHPs (which was not fully supported by the findings); (2) that EAMHPs would be more accurate than laypeople with low levels of horse experience (which was not supported by the findings); and (3) that assessment accuracy would improve as horse experience increases (which was supported by the findings).

The EAAAI Worked

Although this was the first time an instrument such as the EAAAI has been developed and tested, the method used in this study produced significant results. Indeed, participants were able to watch videos of horses and even without any contextual clues or knowledge of the temperaments of the animals, they were able to identify the affective

ACCURACY OF HORSE AFFECT ASSESSMENTS

states of the horses with moderate success, revealing some significant differences in the assessment accuracy scores of the groups.

Just How Accurate Were the Groups?

When we look at the scores obtained by the three groups of participants from the prediction testing phase, we see that overall the scores ranged from -1.19 to +13.31 points across all three groups. NEAMHPs scored between -1.19 and +8.70 points (with an average score of +3.52 points), the laypeople scored between -1.34 and +13.30 points (with an average score of +5.77 points), and the EAMHPs scored between -1.71 and +13.31 points (with an average of +5.12 points).

Compared to the experts' scores presented in the exploratory analysis, the maximum score observed across the three groups (+13.31) appears to be dramatically less than that of the experts (+19.98). This difference of 6.67 points represents a decrease in assessment accuracy of 44% when comparing the best-performing expert participant against the best-performing prediction testing phase participant (who additionally belonged to the EAMHP group). This exploratory analysis allows us to conclude that the participants of the instrument development phase were indeed more accurate than all three of the groups in the prediction testing phase.

As a reminder, the total assessment accuracy score a participant could obtain went as low as -50.0 points. The fact that no participant had a total assessment accuracy score of less than -2 points indicates that people were indeed able to draw some correct conclusions regarding the affective states of horses they will never meet, knew nothing about, and were observing through footage.

Online Delivery

The fact that this study was delivered entirely online is of immense value, as research studies in the field of animal assisted therapy in general are usually limited by their methodological designs (Serpell et al., 2017) which are typically impacted by handler effects (i.e., the effects a handler has on study outcomes due to their necessary presence when a live animal is used in research; Crossman, 2017), small sample sizes (due to time commitments, and the scheduling of participants, researchers, animals, and their handlers), and lack of internal validity (as the animal's behaviour varies from participant to participant). The purely online delivery of this study meant that all these issues were avoided, allowing for an important, flexible, and potentially more reliable way to measure a person's accuracy at assessing equine affect based on video recordings.

Footage as Stimuli

The use of video footage when conducting welfare assessments (through the QBA procedure) is supported by the literature (Clarke et al., 2015; Fleming et al., 2013; Minero 2016, 2018; Wemelsfelder et al., 2001). The complete online delivery of this study (only possible with the use of video footage) not only helped reduce costs and resources and eliminate the methodological design limitations inherent to research with animals, but it also allowed for the study to be dispersed to both Canadians and Americans. In fact, the study's design and the EAAAI developed here for measuring affect assessment accuracy, due to its entirely online nature, could easily be adapted for further research with participants from other countries.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Are EAMHPs More Accurate than NEAMHPs and Laypeople?

Implications of the Findings of the Welch's ANOVA

One of the most interesting findings of this study was the effect profession had on a person's accuracy. When horse experience was not controlled for, both the laypeople and the EAMHPs were more accurate than the NEAMHPs (as demonstrated by the results of the Welch's ANOVA). This observed finding seemed to make sense as the EAMHP group was predominantly composed of participants who had high levels of self-rated horse experience. In addition, the laypeople group was almost evenly split, with nearly half its sample containing participants with none or a low level of horse experience and just over half possessing a medium or high level of self-rated horse experience. In contrast, the NEAMHPs group was predominantly comprised of participants self-identifying as having none or a low level of horse experience.

Regarding NEAMHPs with High Levels of Horse Experience. Interestingly, NEAMHPs with none or a low level of horse experience, who may have been expected to opt out of participation, actually comprised the majority of participants in this group. In fact, the highest horse experience score observed in the NEAMHP group was 42/71, whereas the highest horse experience scores observed in the EAMHP group and laypeople group were 64/71 and 69/71 respectively.

Summary. The finding here that EAMHPs outperformed NEAMHPs initially provided a positive outlook for the use of horses in equine assisted therapy. Indeed, these mental health professionals appeared to be significantly more accurate at assessing the effective states of horses when compared to their non-equine involved counterparts—an encouraging result for those concerned with the ethical use of equines in therapy.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Implications of the Findings of the Multiple Regression Analysis

However, controlling for horse experience scores obtained from participant answers to the horse experience questionnaire eliminated the significant difference in total assessment accuracy scores observed between the EAMHPs and the NEAMHPs. That is, while the results of the multiple regression analysis further corroborated that NEAMHPs are significantly less accurate than laypeople when scored horse experience is controlled for (equal to zero)—the above-described significant difference in total assessment accuracy scores between the NEAMHP group and the EAMHP group is no longer present (although the former still scored lower than the latter). Therefore, neither the prediction that the EAMHPs would be more accurate than the NEAMHPs nor the prediction that EAMHPs would be more accurate than laypeople with low levels of horse experience was supported.

Accuracy Improves as Horse Experience Increases

Previous research has found that veterinary students in training struggle to correctly assess the affective states of horses (Gronqvist et al., 2016). Gronqvist et al. (2016) proposed that inexperienced people are more at risk of injury when working with horses as their lack of experience hinders their ability to interpret and anticipate the behaviours and intentions of the animal. In contrast, then, horse-experienced individuals should be better able to meaningfully interpret equine behaviours, understand the animal's intentions and affective states, and predict how to safely respond to their actions.

The results of the multiple regression in this study offer support for this proposition, as those participants who rated themselves as having a high level of experience were indeed found to be more likely to assess the affective states of the horses

ACCURACY OF HORSE AFFECT ASSESSMENTS

similarly to the experts, compared to those who believed they have no experience or a low level of experience. Further, horse experience scores positively predicted a significant increase in accuracy scores, supporting the theory that the more experience a person has with horses the more likely they are to accurately assess their affective states.

Additionally, the between-subjects one-way ANOVA further demonstrate that participants who rated themselves as having no horse experience or a low level of horse experience were significantly less accurate at assessing the affective states of a horse in comparison to participants who rated themselves as having a high level of horse experience. These results of the multiple regression analysis and the between-subjects one-way ANOVA therefore provide support for the third prediction of this study.

Importance of the Horse Experience Questionnaire

The development of the horse experience questionnaire in this study was undertaken as there are currently no established standardized assessments for horse experience level outside of horse-riding ability or general horse knowledge. In this study, horse experience proved to be an important factor in the performance of the EAMHPs ability to accurately assess equine affect. The questionnaire developed here within offers a fast, easily tallied survey that is both significantly, positively correlated with self-rated levels of horse experience and accounted for a significant level of variance in the observed total accuracy assessment scores. This makes it a valuable tool that would benefit from further development.

The Impact of Horse Experience on Assessment Accuracy

This observed shift in the findings makes it clear that horse experience plays an important role in whether or not EAMHPs are more accurate than NEAMHPs at assessing

ACCURACY OF HORSE AFFECT ASSESSMENTS

equine affective states. When horse experience was controlled for, laypeople were the only ones who significantly outperformed another group (the NEAMHP; a finding that was also observed when horse experience was not controlled for). This is curious as 94.6% of the EAMHPs self-reported having a medium to high level of horse experience (with horse experience scores that ranged from 25–53 points and 23–64 points respectively). In comparison, 42.6% of the laypeople group self-identified as having none to a low level of horse experience (with horse experience scores that ranged from 0–9 points and 2–40 points out of 71, respectively). It seems possible that the EAMHPs possibly overestimated or over-represented their level of horse experience. The question then is: Why do non-horse experienced laypeople still perform significantly better than the NEAMHPs but the EAMHPs, who almost unanimously self-reported having medium to high levels of horse experience did not?

Possible Differences in Types of Horse Interaction. A point of speculation regarding these findings lies in the nature of the relationships horse-experienced laypeople and EAMHPs have with horses. It is possible that a distinction may exist here, where experienced laypeople's involvement with horses may be more direct, casual/private, and animal-focused. In contrast, the contact EAMHPs have with horses may be more likely to take place in less horse-centric work environments, where: (a) the professional's focus may be split between the client and horse; (b) they may be more focused on the client's interpretations of the animal's emotional state for the purposes of meeting therapeutic goals rather than the animal's actual affective state; and/or (c) because of the theoretical model used (e.g., equine-facilitated psychotherapy), the

ACCURACY OF HORSE AFFECT ASSESSMENTS

therapist may not need to be in tune with the emotional state of the horse co-therapist to do their job and therefore may be less skilled at assessing horse affect.

So, while EAMHPs may believe they possess a high level of horse experience (as reflected in the number of participants in this group who rated themselves as such) due to their frequent contact with horses, the nature of this experience may in fact differ in important ways from the experience obtained by laypeople. This in turn may account for the EAMHP group's observed—fairly average—accuracy when assessing horse affective states. If this were found to be true, it would suggest that the horse experience obtained by an EAMHP in the work environment is not in and of itself adequate to permit an EAMHP to accurately assess the affective states of their horse co-therapists. Further research would be needed in order to investigate whether the types of contact horse-involved laypeople and EAMHPs have with horses on a regular basis has any impact on their abilities to assess the affective states of horses with a level of accuracy similar to that of experts.

The State of Equine Welfare in Equine Assisted Therapy Settings

The benefit of equine assisted therapy should not come at the cost of the horse's welfare. Whenever an animal is used for human purposes, it is the human's responsibility to be able to notice and react to changes in the animal's affective state promptly and accurately to ensure the horse is being used ethically and that its mental wellbeing is not being compromised (Mellor, 2016). Failure to do so raises welfare concerns (Gronqvist et al., 2017) and may pose safety risks to both the horse (DuBois et al., 2018; Gronqvist et al., 2016; Horseman et al., 2016; Waran et al., 2007). and the client (Gronqvist et al., 2016).

ACCURACY OF HORSE AFFECT ASSESSMENTS

Therefore, the finding that EAMHPs are not more accurate than laypeople with a low level of horse experience is noteworthy, as misinterpretations of the horse co-therapist's affective state can quickly lead to misunderstandings of behaviour, breakdowns in horse-human communication, and the development of situations that threaten the horse's wellbeing (Gronqvist et al., 2016; see also Horseman et al., 2016; McBride & Long, 2001; Merkies et al., 2018). If a therapist is unable to recognize when a horse's affective state has been compromised, then they may consequently fail to prevent future harm or be unable to identify when to remove the animal from a session in order to safeguard its welfare. Therefore, the average level of accuracy (in comparison to laypeople and NEAMHPs) observed among the EAMHP's may directly impact the welfare of horses used in equine therapy contexts (Merkies et al., 2018).

With the observation that EAMHPs possibly tend to over-estimate their own level of horse experience, and the finding that horse experience makes the difference between an EAMHP being more accurate at assessing equine affect than their non-equine assisted colleagues, the qualifications of those involved in equine assisted therapy are important and clearly there is a need for a means of testing and screening prospective equine assisted therapists, counsellors, and mental health professionals before they begin working with a horse co-therapist. While this is largely relevant if the professional is seeking to work in a modality that utilizes a team of three (comprised of the horse, client, and mental health professional), those professionals also seeking to work in models alongside a horse expert in a team of four would still benefit from having at least some degree of affect assessment accuracy in order to keep themselves and their clients safe. While most of the responsibility to monitor the horse's affective state and behaviour in

ACCURACY OF HORSE AFFECT ASSESSMENTS

some models may fall to a horse experienced professional, rather than to the acting mental health professional, it is arguably still important for the mental health professional to keep their client safe from the horse co-therapist. Therefore, they should possess some degree of horse experience so that they are able to notice large shifts in the animal's emotional state, anticipate potential behaviours that may threaten the client's safety, and then take appropriate actions to prevent that harm from occurring. Further, as the mental health professional has chosen to work in an equine assisted setting, they are also responsible for the ethical use and employment of that animal, as noted in the Canadian Code of Ethics for Psychologists (CPA, 2017). Therefore, having the ability to assess the affective state of the horse co-therapist with some degree of accuracy should still be the responsibility of all mental health professionals who employ horses in their practice, regardless of the model they adhere to.

In this study, we have introduced the Equine Affect Assessment Accuracy Indicator. The use of the EAAAI involves the completion of the horse experience questionnaire, the presentation of ten video recordings of horses, the process of having participants generate 1–10 expressive terms of affect and their associated descriptions, and using the code developed here to categorize and score those terms.

Anthropomorphic Descriptions of Affect

The use of anthropomorphism is a strength of this study, as without it, there would exist no way to non-invasively assess the affective states of horses and it would have been impossible to develop an instrument to measure affect assessment accuracy. Frequently, the use of integrative animal assessments and descriptive terms to summarize an animal's demeanour has been dismissed due to their anthropomorphic undertones (Wemelsfelder

ACCURACY OF HORSE AFFECT ASSESSMENTS

et al., 2000). Yet, by looking at animals through anthropomorphic eyes and describing what we observe, humans not only came to domesticate animals as pets and work companions, but this also led to the development of animal welfare and proper husbandry practices (Mithen, 1996).

Research utilizing the QBA procedure has demonstrated that anthropomorphic language and terms can be scientifically analysed (Wemelsfelder et al., 2000). Further, their use in integrative welfare measurements (such as the QBA) is backed by consistently high inter-observer reliability findings (Wemelsfelder et al., 2000). Therefore, descriptive terms of demeanour and the use of integrative welfare measurements should not be condemned as anthropomorphic folly (Wemelsfelder et al., 2000). It is inappropriate to dismiss the use of the QBA or the generation of affective terms based on their seemingly anthropomorphic nature as years of research utilizing the QBA have confirmed that human observation and perception, as an assessment tool, is reliable (Wemelsfelder et al., 2000).

Further, the very techniques used in animal assisted therapy and equine assisted therapy tend to already be anthropomorphic in nature, as they uniquely offer therapeutic value by way of the human client's interpretations of spontaneous interactions with the animal co-therapist. If it is appropriate to consider such anthropomorphic interpretations in animal assisted therapy, then it is also acceptable, given the plethora of research findings, to analyze participant-generated descriptive terms of affect as well. The use of expert participants' descriptive terms of affect to create an answer key in the instrument development phase allowed for a method (and tool—the EAAAI) through which to assess

ACCURACY OF HORSE AFFECT ASSESSMENTS

accuracy to be introduced by this study, a task which otherwise might have been impossible to achieve.

Sample Gender Distribution

Further, the demographics of the instrument development phase and prediction testing phase samples were determined to be representative of the populations of interest which adds merit to the findings presented here. While the samples of both phases were predominantly female this distribution was unsurprising, as females outnumber males both in the Canadian (Evans, 2010) and American (Stowe, 2018) equine industry. Female participation also tends to outnumber male participation in equine welfare-related research in Canada (DuBois et al., 2018; Merkies et al., 2017), and the United States (Lofgren et al., 2016). Further, Canadian (Khanam et al., 2022) and American (American Psychological Association, 2011; Salsberg et al., 2020; U.S. Bureau of Labor Statistics, 2020) mental health professions are primarily female-dominated.

The predominantly female-identifying samples recruited in this study were therefore not deemed to be problematic and are likely representative of the populations of interest. However, the generalizability of the findings of this study may be limited if non-female identifying participants are tested using the EAAAI.

Limitations

Categorization of Terms

This study does contain some other limitations that must be considered. First, as the sole assessor of the terms and their categorization in this study, the findings herein may have been impacted. To mitigate potential bias, the rules in Appendix E were prepared prior to beginning any coding processes or interpretations. They were then

ACCURACY OF HORSE AFFECT ASSESSMENTS

followed accordingly, alongside the use of the aforementioned resources (i.e., the dictionary, thesaurus, and feeling wheel), to inform categorization decisions.

Possible Impact of Self-Selection Among Laypeople

Second, self-selection bias amongst the layperson group may help account for the finding that only the laypeople significantly perform better than the NEAMHPs, when horse experience is accounted for, but the EAMHPs do not. As participants from the layperson group were entirely recruited via self-selection it is also possible that people with more interest in and involvement with horses were more likely to dedicate time to fully completing the study, which could have impacted the findings. This use of self-selection is a limitation of this study, and the online modality of its delivery may have furthered its impact (Heiervang & Goodman, 2011). It is worth noting that potential NEAMHP participants were not screened for their personal involvement with horses. Only whether they made mention of offering equine assisted therapy related services on their website or not impacted their eligibility for recruitment into the NEAMHP group.

Challenges Related to Participant Recruitment

Third, the collection of participants for both phases of this study posed an immense challenge. Ideally, the first phase of the study would have had a much larger sample size. However, to the best of my knowledge, every eligible certified animal behavioural consultant (who works with horses) in Canada was approached for the study and even after extending into the United States, it took six months to recruit the four participants needed from this profession in the instrument development phase of the study. In the prediction testing phase, 56/386 and 51/745 potential EAMHP and NEAMHP participants contacted over the course of six months completed the study.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Timeline for Equine Assisted Therapy Involvement When Screening for EAMHPs

Fourth, it would have been beneficial to the study if a timeline had been specified in the question “Are horses involved in your practice?” which was used to help categorize participants as either NEAMHPs or EAMHPs. In one case a participant indicated that they had previously worked as an equine assisted therapist but had stopped a year prior to the testing date. For the purposes of this study that individual was categorized as a NEAMHP, but it is impossible to know if other participants were potentially sorted as EAMHPs incorrectly in this way.

Recommendations

Given the findings of this study, further research is recommended to continue the investigation into the welfare of horses being used in equine assisted therapy, and into the handling/husbandry practices being employed by mental health professionals who use equine assisted therapy models. To this end, the most appropriate next steps are outlined below.

Considerations for the Training of EAMHPs

Based on the preliminary findings of this study, EAMHPs should receive training in recognizing and interpreting horse behaviour, and they should possess considerable horse experience prior to practicing EAT (especially if they do not intend to operate out of a model that utilizes a 4th team member who exclusively attends to the horse’s behaviour and needs). It is recommended, based on the findings of this study that such training would also benefit from including the development of the awareness that not all horse experience may be equal, and that EAMHPs may inadvertently overestimate their

ACCURACY OF HORSE AFFECT ASSESSMENTS

ability to understand and account for the horses needs based on their position as an EAMHP even if they lack the experience and knowledge required to do so.

Anchoring Accuracy with New Experts

The exploratory analysis conducted in this study demonstrated that the participants of the instrument development phase were more accurate than all three of the prediction testing phase groups. This in turn, supports the use of the experts in the development of the EAAAI as they are clearly more accurate at assessing horse affect than the prediction testing phase participants. However, we cannot conclude for certain how accurate a person must be before they are considered “good enough” at assessing equine affect so as not to cause harm.

It would be beneficial for another phase of the study to be conducted involving a new set of expert participants (of the same professional makeup as conducted in the instrument development phase of this study) to see how well they perform. This would be an important next step to take, as it would allow for the total assessment accuracy scores to be meaningfully anchored. It may also permit for conclusions to be drawn regarding the above question: “how much accuracy is enough” and it would further allow for the contents of the categories to be expanded.

Any recruitment of future animal behavioural consultants (who work with horses) should be done outside of Canada to avoid retesting, as this population was exhausted during recruitment for the instrument development phase of this study.

Reducing Classification Bias with a Hybrid Term Generation Methodology

Should another phase of this study be completed in the future, a hybrid FCP and fixed list methodology may be useful to reduce potential bias during the classification

ACCURACY OF HORSE AFFECT ASSESSMENTS

process. FCP should be primarily used to preserve the nuances of participant assessments; however, the study activity should be programmed in such a way that if a participant entered a term listed here as needing further classification (see Appendix E), a list of the associated classifiers would then subsequently appear, prompting the participant to clarify their use of the term. Thus, any bias introduced by the author's interpretation would be eliminated for such terms.

It is important to reiterate that a fixed list was not used in this study as it was crucial to avoid influencing the participant's observations and to avoid prompting an otherwise ignorant participant in such a way that they appear knowledgeable. This paper focuses on the ethical use of, and welfare impacts on, horses in therapy settings; therefore, the nuances of emotion and a person's ability to recognize those subtleties were of primary concern.

Regarding Self-Rated Levels of Horse Experience

Given the current dataset, the overlapping of horse experience scores observed between the self-rated levels of horse experience made it impossible to distinguish clear cut-offs regarding what scores constitute a low, medium, and high level of experience (whereby, arguably no experience should be linked to a horse experience score of zero) at this time. Due to this overlapping of horse experience scores across levels, it seemed likely that participants would have been incorrectly clumped together with other people who possess notably different degrees of actual horse experience, had the variable of self-rated level of horse experience been used in the examination of the effect horse experience has on assessment accuracy.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Therefore, it was determined that further use (specifically in the multiple regression analysis) of the variable of self-rated level of horse experience in this study, based on the present dataset, was likely to wipe out any potential effects other than those of the poles and as a result, only the horse experience scores calculated from the horse experience questionnaire were examined further. Future research would benefit from continuing to ask participants to self-rate their level of horse experience, alongside completion of the horse experience questionnaire to see if a greater amount of data would help reveal natural cut-offs for what constitutes a low, medium, and high level of horse experience (in the context of horse experience questionnaire score ranges).

Purposes the EAAAI Could be Adapted For

There is a need for the continued development of psychometrically rigorous methods for assessing competencies relating to working with horse. With further testing, data collection, and revising, the EAAAI could be adapted in the future with the aim of using it to: screen prospective therapists looking to enter the equine assisted therapy field; sorting participants into the most appropriate courses/levels (e.g., beginner, intermediate, advanced) for equine therapy training programs; or it could also be used as an examination of readiness/competency to be delivered to prospective mental health professionals graduating from equine assisted therapy focused classes/programs. Such competency testing could represent a first, tentative step towards the implementation of regulations to protect the ethical use of horse co-therapists in equine assisted therapy settings.

Conclusion

The current study, the EAAAI, and its findings are worth careful consideration as they represent the first attempt in the field of equine assisted therapy research to assess the accuracy of EAMHPs on their abilities to interpret the emotional states of their equine co-therapists. The methodology used within has laid the foundation for the continuation of future research to investigate the ethical use of horses in equine assisted therapy contexts. Further, the EAAAI developed here could potentially be adapted for the testing of prospective EAMHPs in the future.

This study set out to investigate (a) whether EAMHPs are more accurate (in the context of this study) than NEAMHPs and horse experienced laypeople when it comes to assessing equine affective states similarly to the experts, and (b) if having higher levels of horse experience impacts this assessment accuracy. The finding that EAMHPs do not score significantly different from NEAMHPs and laypeople when horse experience is accounted for is concerning and may bear ethical implications for the unregulated use of horses in equine assisted therapy contexts in Canada and the United States. More research needs to be conducted to further investigate the involvement of horses in equine assisted therapy practices to ensure that their use remains ethical and that their welfare is not being compromised in exchange for aiding clients.

References

- American Horse Council Foundation. (2018). *Economic impact of the U.S. horse industry*. American Horse Council.
- American Psychological Association. (2011). *Data tool: Demographics of the U.S. psychology workforce*. Retrieved July 27, 2023, from <https://www.apa.org/workforce/data-tools/demographics>
- Arena, L., Wemelsfelder, F., Messori, S., Ferri, N., & Barnard, S. (2017). Application of free choice profiling to assess the emotional state of dogs housed in shelter environments. *Applied Animal Behaviour Science*, 195, 72–79. <https://doi.org/10.1016/j.applanim.2017.06.005>
- Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders*, 39(9), 1261–1267. <https://doi:10.1007/s10803-009-0734-3>
- Bona, E. (2020). *Caring with critters: An exploration workshop in animal-assisted interventions*. Dreamcatcher: An animal assisted wellness academy.
- Canadian Psychological Association - CPA. (2017). *Canadian code of ethics for psychologists* (4th ed.) https://cpa.ca/docs/File/Ethics/CPA_Code_2017_4thEd.pdf
- Clarke, T., Pluske, J. R., & Fleming, P. A. (2016). Are observer ratings influenced by prescription? A comparison of free choice profiling and fixed list methods of qualitative behavioural assessment. *Applied Animal Behaviour Science*, 177, 77–83. <https://doi.org/10.1016/j.applanim.2016.01.022>
- Coffey, M. (2013, May 13). *Dogs helping humans heal*. Retrieved from http://www.youtube.com/watch?v=U_nDaGzam0A
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.
- Coren, S. (2015). Foreword. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (4th ed., pp. xix–xxii). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-801292-5.06001-6>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Coren, S. (2019). Foreword. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed., pp. xxv–xxvii). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-815395-6.06001-6>
- Crossman, M. K. (2017). Effects of interactions with animals on human psychological distress. *Journal of Clinical Psychology*, 73(7), 761–784. <https://doi.org/10.1002/jclp.22410>
- Dijksterhuis, G. B., & Gower, J. C. (1991). The interpretation of generalized Procrustes analysis and allied methods. *Food Quality and Preference*, 3(2), 67–87. [https://doi:10.1016/0950-3293\(91\)90027-c](https://doi:10.1016/0950-3293(91)90027-c)
- DuBois, C., Nakonechny, L., Derisoud, E., & Merkies, K. (2018). Examining Canadian equine industry participants' perceptions of horses and their welfare. *Animals*, 8(11), 201. <https://doi:10.3390/ani8110201>
- EAGALA. (n.d.). *About*. Retrieved November 15, 2020, from <https://www.eagala.org/org>
- Evans, V. (2010). *Canadian horse industry profile study*. Equine Canada. Retrieved July 27, 2023, from <https://www.equestrian.ca/industry/about>
- Farm Animal Welfare Council. (2009, October). *Farm animal welfare in Great Britain: Past, present and future*. GOV.UK. <https://www.gov.uk/government/publications/fawc-report-on-farm-animal-welfare-in-great-britain-past-present-and-future>
- Farm Animal Welfare Council. (2012, October 9). *Five freedoms*. The National Archives. <https://webarchive.nationalarchives.gov.uk/ukgwa/20121010012427/http://www.fawc.org.uk/freedoms.htm>
- Fédération Equestre Internationale. (2020, September 12). *Equestrian nation: Canada*. Retrieved June 30, 2023, from <https://www.fei.org/stories/lifestyle/my-equestrian-life/equestrian-nation-canada#:~:text=A%20survey%20conducted%20in%20recent,using%20horses%20owned%20by%20others>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Fine, A. H., & Beck, A. M. (2019). Chapter 1 – Understanding our kinship with animals: input for health care professionals interested in the human–animal bond. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed., pp. 3–12). Elsevier Inc.
<https://doi.org/10.1016/B978-0-12-815395-6.00001-8>
- Fleming, P. A., Paisley, C. L., Barnes, A. L., & Wemelsfelder, F. (2013). Application of qualitative behavioural assessment to horses during an endurance ride. *Applied Animal Behaviour Science*, *144*(1), 80–88.
<https://doi.org/10.1016/j.applanim.2012.12.001>
- Fleming, P. A., Wickham, S. L., Stockman, C. A., Verbeek, E., Matthews, L., & Wemelsfelder, F. (2015). The sensitivity of QBA assessments of sheep behavioural expression to variations in visual or verbal information provided to observers. *Animal*, *9*(5), 878–887. <https://doi.org/10.1017/S1751731114003164>
- Frenería y Romanas López. (n.d.). *Nosebands*. Retrieved June 26, 2023, from http://www.frenerialopez.com/cat3/index.php?dispatch=categories.view&category_id=4
- Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, *40*(1), 33–51.
<https://doi:10.1007/bf02291478>
- Granados, A. C., & Agís, I. F. (2011). Why children with special needs feel better with hippotherapy sessions: A conceptual review. *Journal of Alternative and Complementary Medicine*, *17*(3), 191–197.
<https://doi.org/10.1089/acm.2009.0229>
- Grier, J. W. (1984). *Biology of Animal Behavior*. Times Mirror/Mosby.
- Gronqvist, G., Rogers, C., Gee, E., Bolwell, C., & Gordon, S. (2016). The challenges of using horses for practical teaching purposes in veterinary programmes. *Animals*, *6*(11), 69. <https://doi.org/10.3390/ani6110069>
- Gronqvist, G., Rogers, C., Gee, E., Martinez, A., & Bolwell, C. (2017). Veterinary and equine science students' interpretation of horse behaviour. *Animals*, *7*(8), 63.
<https://doi.org/10.3390/ani7080063>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Hall, C., & Heleski, C. (2017). The role of the ethogram in equitation science. *Applied Animal Behaviour Science*, 119, 102–110. <https://doi.org/10.1016/J.APPLANIM.2017.02.013>
- Hall, C., Goodwin, D., Heleski, C., Randle, H., & Waran, N. (2008). Is there evidence of learned helplessness in horses? *Journal of Applied Animal Welfare Science: JAAWS*, 11, 249–266. <https://doi.org/10.1080/10888700802101130>
- Hausberger, M., Muller, C., & Lunel, C. (2011). Does work affect personality? A study in horses. *PloS One*, 6, e14659. <https://doi.org/10.1371/journal.pone.0014659>
- Heidenreich, B. (2012). An introduction to the application of science-based training technology. *Veterinary Clinics of North America: Exotic Animal Practice*, 15(3), 371–385. <https://doi.org/10.1016/j.cvex.2012.06.006>
- Heievang, E., & Goodman, R. (2011). Advantages and limitations of web-based surveys: Evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, 46(1), 69–76. <https://doi:10.1007/s00127-009-0171-9>
- Hintze, S., Murphy, E., Bachmann, I., Wemelsfelder, F., & Würbel, H. (2017). Qualitative behaviour assessment of horses exposed to short-term emotional treatments. *Applied Animal Behaviour Science*, 196, 44–51. <https://doi.org/10.1016/j.applanim.2017.06.012>
- Horseman, S. V., Buller, H., Mullan, S., & Whay, H. R. (2016). Current welfare problems facing horses in Great Britain as identified by equine stakeholders. *PLOS ONE*, 11(8), e0160269. <https://doi.org/10.1371/journal.pone.0160269>
- Horses of the Sun. (n.d.). *Voltige*. Retrieved June 26, 2023, from <https://horsesofthesun.weebly.com/voltige.html>
- Human-Equine Alliance for Learning. (n.d.). *Equine therapy types*. <https://humanequinealliance.com/the-heal-model/equine-therapy-types/>
- Jenkins, C. D., Laux, J. M., Ritchie, M. H., & Tucker-Gail, K. (2014). Animal-assisted therapy and Rogers' core components among middle school students receiving counseling services: A descriptive study. *Journal of Creativity in Mental Health*, 9(2), 174–197. <http://dx.doi.org/10.1080/15401383.2014.899939>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Jørgensen, G. H. M., Liestøl, S. H. O., & Bøe, K. E. (2011). Effects of enrichment items on activity and social interactions in domestic horses (*Equus caballus*). *Applied Animal Behaviour Science*, *129*(2–4), 100–110. <https://doi.org/10.1016/j.applanim.2010.11.004>
- Kaiser, L., Heleski, C., Siegford, J., & Smith, K. (2006). Stress-related behaviors among horses used in a therapeutic riding program. *Journal of the American Veterinary Medical Association*, *228*, 39–45. <https://doi.org/10.2460/javma.228.1.39>
- Khanam, F., Langevin, M., Savage, K., & Uppal, S. (2022). *Insights on Canadian society: Women working in paid care occupations*. Statistics Canada. Retrieved July 27, 2023, from <https://www150.statcan.gc.ca/n1/pub/75-006-x/2022001/article/00001-eng.htm>
- Krüger, K. (2007). Behaviour of horses in the “round pen technique.” *Applied Animal Behaviour Science*, *104*(1-2), 162–170. <https://doi.org/10.1016/j.applanim.2006.04.021>
- Lac V., & Walton R. (2012). Companion animals as assistant therapists: Embodying our animal selves. *British Gestalt Journal*, *21*(1), 32–39. <https://doi.org/10.53667/SKSC8214>
- Latella, D., & Abrams, B. N. (2019). Chapter 10: The role of the equine in animal-assisted interactions. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed., pp. 133–162). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-815395-6.00010-9>
- Levinson, B. M. (1978). Pets and personality development. *Psychological Reports*, *42*(3_suppl), 1031–1038. <https://doi.org/10.2466/pr0.1978.42.3c.1031>
- Levinson, B. M., & Mallon, G. P. (1997). *Pet-oriented child psychotherapy* (2nd ed.). Charles C. Thomas.
- Lofgren, E. A., Voigt, M. A., & Brady, C. M. (2016). Information-seeking behavior of the horse competition industry: A demographic study. *Journal of Equine Veterinary Science*, *37*, 58–62. <https://doi.org/10.1016/j.jevs.2015.10.005>
- Masłowska, K., Mizzoni, F., Dwyer, C. M., & Wemelsfelder, F. (2020). Qualitative behavioural assessment of pain in castrated lambs. *Applied Animal Behaviour Science*, *233*, 105–143. <https://doi.org/10.1016/j.applanim.2020.105143>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Mauricio, A. A., Palazzo, A. B., Caselato, V. M., and Bolini, H. M. A. (2016). Generalized Procrustes analysis and external preference map used to consumer drivers of diet gluten free product. *Food and Nutrition Sciences*, 7, 711–723. <http://dx.doi.org/10.4236/fns.2016.79072>
- Mcbride, S., & Long, L. (2001). Management of horses showing stereotypic behaviour, owner perception and the implications for welfare. *The Veterinary Record*, 148, 799–802. <https://doi.org/10.1136/vr.148.26.799>
- McDonnell, S. M., & Haviland, J. C. S. (1995). Agonistic ethogram of the equid bachelor band. *Applied Animal Behaviour Science*, 43(3), 147–188. [https://doi.org/10.1016/0168-1591\(94\)00550-X](https://doi.org/10.1016/0168-1591(94)00550-X)
- McGreevy, P. D. (2004) *Equine Behaviour* (2nd ed.). Saunders Ltd.
- McGreevy, P. D., & McLean, A. N. (2009). Punishment in horse-training and the concept of ethical equitation. *Journal of Veterinary Behavior*, 4(5), 193–197. <https://doi.org/10.1016/j.jveb.2008.08.001>
- Mellor, D. J. (2014). Positive animal welfare states and reference standards for welfare assessment. *New Zealand Veterinary Journal*, 63(1), 17–23. <https://doi.org/10.1080/00480169.2014.926802>
- Mellor, D. J. (2016). Updating animal welfare thinking: Moving beyond the “five freedoms” towards “a life worth living.” *Animals*, 6(3), 21. <https://doi.org/10.3390/ani6030021>
- Mellor, D. J., & Beausoleil, N. (2015). Extending the “five domains” model for animal welfare assessment to incorporate positive welfare states. *Animal Welfare*, 24. <https://doi.org/10.7120/09627286.24.3.241>
- Mellor, D. J., & Reid, C. S. W. (1994). Concepts of animal well-being and predicting the impact of procedures on experimental animals. *Improving the well-being of animals in the research environment*, 3–18. <https://www.wellbeingintlstudiesrepository.org/exprawel/7/>
- Merkies, K., Crouchman, E., & Belliveau, H. (2022). Human ability to determine affective states in domestic horse whinnies. *Anthrozoos*, 35(3), 483–494. <https://doi.org/10.1080/08927936.2021.1999605>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Merkies, K., Mckechnie, M., & Zakrajsek, E. (2018). Behavioural and physiological responses of therapy horses to mentally traumatized humans. *Applied Animal Behaviour Science*. <https://doi.org/10.1016/j.applanim.2018.05.019>
- Merkies, K., Nakonechny, L., DuBois, C., & Derisoud, E. (2017). Preliminary study on current perceptions and usage of training equipment by horse enthusiasts in Canada. *Journal of Applied Animal Welfare Science*, 21(2), 141–152. <https://doi.org/10.1080/10888705.2017.1392301>
- Merriam-Webster. (2023a). *Hopeful*. <https://www.merriam-webster.com/dictionary/hopeful>
- Merriam-Webster. (2023b). *Joy*. <https://www.merriam-webster.com/dictionary/joy>
- Merriam-Webster. (n.d.-a). *Affect*. <https://www.merriam-webster.com/dictionary/affect>
- Merriam-Webster. (n.d.-b). *Dictionary*. <https://www.merriam-webster.com/dictionary>
- Merriam-Webster. (n.d.-c). *Thesaurus*. <https://www.merriam-webster.com/thesaurus>
- Minero, M., & Canali, E. (2009). Welfare issue of horses: An overview and practical recommendations. *Italian Journal of Animal Science*, 8, 219–230. <https://doi.org/10.4081/ijas.2009.s1.219>
- Minero, M., Dalla Costa, E., Dai, F., Canali, E., Barbieri, S., Zanella, A., Pascuzzo, R., & Wemelsfelder, F. (2018). Using qualitative behaviour assessment (QBA) to explore the emotional state of horses and its association with human-animal relationship. *Applied Animal Behaviour Science*, 204, 53–59. <https://doi.org/10.1016/j.applanim.2018.04.008>
- Minero, M., Dalla Costa, E., Dai, F., Murray, L. A. M., Canali, E., & Wemelsfelder, F. (2016). Use of qualitative behaviour assessment as an indicator of welfare in donkeys. *Applied Animal Behaviour Science*, 174, 147–153. <https://doi.org/10.1016/j.applanim.2015.10.010>
- Minero, M., Tosi, M. V., Canali, E., & Wemelsfelder, F. (2009). Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. *Applied Animal Behaviour Science*, 116(1), 74–81. <https://doi.org/10.1016/j.applanim.2008.07.001>

ACCURACY OF HORSE AFFECT ASSESSMENTS

Mithen, S. (1996). *The prehistory of the mind: An exchange*. Thames and Hudson.

Napolitano, F., De Rosa, G., Braghieri, A., Grasso, F., Bordi, A., Wemelsfelder, F. (2008). The qualitative assessment of responsiveness to environmental challenge in horses and ponies. *Applied Animal Behaviour Science* 109, 342–354. <https://10.1016/j.applanim.2007.03.009>

Nilson, R. (2004). Equine-facilitated psychotherapy. *Perspectives in Psychiatric Care*, 40(2), <https://doi.org/10.1111/j.1744-6163.2004.00042.x>

PATH International. (n.d.-a), *About PATH Intl*. Retrieved July 4, 2023, from <https://pathintl.org/about/>

PATH International. (n.d.-b), *Therapists and equine-assisted services*. Retrieved July 4, 2023, from <https://pathintl.org/programs/therapies-incorporating-equines/>

Reega, S. J. (2017). Effects of equine assisted activities and therapies on equine stress and welfare. [Senior Honours Thesis, University of New Hampshire, Durham]. Honors Theses and Capstones. <https://scholars.unh.edu/honors/369>

Ringhofer, M., & Yamamoto, S. (2017). Domestic horses send signals to humans when they face with an unsolvable task. *Animal Cognition*, 20(3), 397–405. <https://doi.org/10.1007/s10071-016-1056-4>

Rousing, T., & Wemelsfelder, F. (2006). Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Applied Animal Behaviour Science*, 101(1), 40-53. <https://doi.org/10.1016/j.applanim.2005.12.009>

Rutherford, K. M. D., Donald, R. D., Lawrence, A. B., & Wemelsfelder, F. (2012). Qualitative behavioural assessment of emotionality in pigs. *Applied Animal Behaviour Science*, 139(3–4), 218–224. <https://doi.org/10.1016/j.applanim.2012.04.004>

Salsberg, E., Quigley, L., Richwine, C., Sliwa, S., Acquaviva, K., & Wyche, K. (2020, August). *The social work profession: Findings from three years of surveys of new social workers*. Fitzhugh Mullan Institute for Health Workforce Equity. <https://www.cswe.org/CSWE/media/Workforce-Study/The-Social-Work-Profession-Findings-from-Three-Years-of-Surveys-of-New-Social-Workers-Dec-2020.pdf>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Scheumann, M., Hasting, A. S., Kotz, S. A., & Zimmermann, E. (2014). The voice of emotion across species: How do human listeners recognize animals' affective states? *PLoS ONE*, *9*(3), e91192. <https://doi.org/10.1371/journal.pone.0091192>
- Serpell, J., McCune, S., Gee, N., & Griffin, J. A. (2017). Current challenges to research on animal-assisted interventions. *Applied Developmental Science*, *21*(3), 223–233. <https://doi.org/10.1080/10888691.2016.1262775>
- Serpell, J. A. (2019). Chapter 2: Animal-assisted interventions in historical perspective. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed., pp. 13–22). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-815395-6.00002-X>
- Shubert, J. (2012). Therapy dogs and stress management assistance during disasters. *U.S. Army Medical Department Journal*, 74–78.
- Snider, L., Korner-Bitensky, N., Kammann, C., Warner, S., & Saleh, M. (2007). Horseback riding as therapy for children with cerebral palsy: is there evidence of its effectiveness? *Physical & Occupational Therapy in Pediatrics*, *27*(2), 5–23. https://doi.org/10.1300/J006v27n02_02
- SPCA. (n.d.). *Five freedoms vs five domains*. Retrieved July 4, 2023, from <https://www.spcacertified.nz/learn-more/article/five-freedoms-vs-five-domains?cat=five-domains&subcat=>
- Statistics Canada. (2023, June 30). *Canada's population clock (real-time model)*. <https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2018005-eng.htm>
- Stockman, C. A., Collins, T., Barnes, A. L., Miller, D., Wickham, S. L., Beatty, D. T., Blache, D., Wemelsfelder, F., & Fleming, P. A. (2013). Flooring and driving conditions during road transport influence the behavioural expression of cattle. *Applied Animal Behaviour Science*, *143*(1), 18–30. <https://doi.org/10.1016/j.applanim.2012.11.003>
- Stockman, C. A., McGilchrist, P., Collins, T., Barnes, A. L., Miller, D., Wickham, S. L., Greenwood, P. L., Cafe, L. M., Blache, D., Wemelsfelder, F., & Fleming, P. A. (2012). Qualitative behavioural assessment of angus steers during pre-slaughter handling and relationship with temperament and physiological responses. *Applied Animal Behaviour Science*, *142*(3–4), 125–133. <https://doi.org/10.1016/j.applanim.2012.10.016>

ACCURACY OF HORSE AFFECT ASSESSMENTS

Stowe, J., C. (2018). *2018 American Horse Publications (AHP) equine industry survey sponsored by Zoetis*. American Horse Publications. Retrieved July 27, 2023, from https://www.americanhorsepubs.org/wp-content/uploads/2018/07/AHP_FinalReport_2018-for-website.pdf

The Humane Society of the United States. (n.d.). *What is soring? Important facts about this cruel abuse*. Retrieved Nov 15, 2020, from <https://www.humanesociety.org/resources/what-soring>

Thesaurus.com. (n.d.). *Thesaurus*. <https://www.thesaurus.com>

Torcivia, C., & McDonnell, S. (2021). Equine discomfort ethogram. *Animals: An open access journal from MDPI*, 11(2), 580. <https://doi.org/10.3390/ani11020580>

Trotter, K. S., Chandler, C. K., Goodwin-Bond, D., & Casey, J. (2008). A comparative study of the efficacy of group equine assisted counseling with at-risk children and adolescents. *Journal of Creativity in Mental Health*, 3(3), 254–284. <https://doi:10.1080/15401380802356880>

U.S. Bureau of Labor Statistics. (2020). *Labour force statistics from the current population survey*. <https://www.bls.gov/cps/aa2020/cpsaat11.htm>

United States Census Bureau. (2023, June 30). *U.S and world population clock*. <https://www.census.gov/popclock/>

VanFleet, R., Fine, A. H., & Faa-Thompson, T. (2019). Chapter 15: Applications of animal-assisted interventions in professional mental health settings: An overview of practice considerations. In A. H. Fine (Ed.), *Handbook on animal-assisted therapy: Theoretical foundations and guidelines for practice* (5th ed., pp. 225–248). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-815395-6.00015-8>

Vindevoghel, T. V., Fleming, P. A., Hyndman, T. H., Musk, G. C., Laurence, M., & Collins, T. (2019). Qualitative behavioural assessment of bos indicus cattle after surgical castration. *Applied Animal Behaviour Science*, 211, 95–102. <https://doi.org/10.1016/j.applanim.2018.11.004>

Visser, E. K., VanDierendonck, M., Ellis, A. D., Rijksen, C., & Van Reenen, C. G. (2009). A comparison of sympathetic and conventional training methods on responses to initial horse training. *The Veterinary Journal*, 181(1), 48–52. <https://doi.org/10.1016/j.tvjl.2009.03.009>

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Waite, C., & Bourke, L. (2013). It's different with a horse. *Youth Studies Australia*, 32(4), 15–24.
- Walker, J. K., Dale, A. R., D'Eath, R. B., & Wemelsfelder, F. (2009). Welfare assessment in dogs: Reducing the invasiveness of welfare assessment methodology through trial and validation of a qualitative, behavioral-based approach. *Journal of Veterinary Behavior*, 4(2), 102–104. <https://doi.org/10.1016/j.jveb.2008.10.032>
- Waran, N., McGreevy, P., & Casey, R. A. (2007). Chapter 7 – Training methods and horse welfare. In N. Waran (Ed.), *The welfare of horses* (pp. 151–180). Springer. https://doi:10.1007/978-0-306-48215-1_7
- Webster, J. (1994). *Animal welfare: A cool eye towards Eden*. Blackwell Science.
- Webster, J. (2008). *Animal welfare: Limping towards Eden*. Wiley.
- Wemelsfelder, F., Hunter, E. A., Mendl, M. T., & Lawrence, A. B. (2000). The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science*, 67(3), 193–215. [https://doi.org/10.1016/S0168-1591\(99\)00093-3](https://doi.org/10.1016/S0168-1591(99)00093-3)
- Wemelsfelder, F., Hunter, E. A., Mendl, M. T., & Lawrence, A. B. (2001). Assessing the 'whole animal': A free choice profiling approach. *Animal Behaviour*, 62(2), 209–220. <https://doi.org/10.1006/anbe.2001.1741>
- Wickham, S. L., Collins, T., Barnes, A. L., Miller, D. W., Beatty, D. T., Stockman, C. A., Blache, D., Wemelsfelder, F., & Fleming, P. A. (2015). Validating the use of qualitative behavioral assessment as a measure of the welfare of sheep during transport. *Journal of Applied Animal Welfare Science*, 18(3), 269–286. <https://doi.org/10.1080/10888705.2015.1005302>
- Willcox, G. (2020). *The feeling wheel, version 2*. The Gottman Institute. https://cdn.gottman.com/wp-content/uploads/2020/12/The-Gottman-Institute_The-Feeling-Wheel_v2.pdf
- Yarnell, K., Hall, C., Royle, C., & Walker, S. (2015). Domesticated horses differ in their behavioural and physiological responses to isolated and group housing. *Physiology & Behavior*. <https://doi.org/10.1016/j.physbeh.2015.02.040>

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 1

Table 1

*Instrument Development Phase: Results of the Generalized Procrustes Analysis:
Residuals by Object*

Object	Residuals	Object	Residuals
Video A	36753.312	Video K	26775.390
Video B	38724.366	Video L	23183.969
Video C	47385.203	Video M	34319.341
Video D	21344.337	Video N	46596.500
Video E	38066.964	Video O	16833.301
Video F	26474.178	Video P	28332.415
Video G	28303.945	Video Q	30180.060
Video H	22086.361	Video R	29669.900
Video I	26069.511	Video S	39525.280
Video J	36453.879	Video T	40109.808

Note. Typically, those objects with the smallest number of residuals have the greater degree of consensus among the observers. The range of residuals presented here are reflective of variation among the experts regarding the number of terms each one produced per video. The fact that they begin in the thousands therefore does not invalidate the findings.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 2

Table 2

Descriptive Analysis of the Prediction Testing Phase Participant Sample

Demographic characteristic	Layperson		EAMHP ^a		NEAMHP ^b	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Age						
18-24	32	34	1	1.8	-	-
25-34	14	14.9	13	23.2	14	27.5
35-44	15	16	19	33.9	15	29.4
45-54	10	10.64	11	19.6	8	15.7
55-64	18	19.15	9	17.9	9	17.6
65-74	5	5.31	2	3.6	5	9.8
75+ ^c	-	-	-	-	-	-
Gender						
Female	82	87.23	51	92.9	38	74.5
Male	10	10.64	3	5.4	11	21.6
Non-binary / Other	2	2.12	-	-	2	3.9
Prefer not to Say	-	-	1	1.8	-	-
Self-rated Level of Horse Experience						
None	17	18.1	-	-	20	39.2
Low	23	24.5	2	5.4	25	49
Medium	20	21.3	18	32.1	4	7.8
High	34	36.2	35	62.5	2	3.9

Note. *N* = 200 participants (consisting of 94 laypersons, 55 EAMHPs, and 51 NEAMHPs).

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

^c Zero participants reported being 75 or older

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 3

Table 3

Correlations Related to Horse Experience Scores and Self-rated Levels of Horse Experience

Variable	1	2
Overall ^a		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.742**	—
Layperson ^b		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.703**	—
EAMHP ^c		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.533**	—
NEAMHP ^d		
1. Scored Horse Experience	—	
2. Self-Rated Level of Horse Experience	.647**	—

^a A Kendall's tau-b correlation determined that the relationship between the prediction testing phase participants' ($n = 200$) self-rated levels of horse experience and the horse experience scores they received had a statistically significant, strong, positive association ($\tau_b = .742, p < .001$).

^b After splitting the data by profession, a second Kendall's tau-b correlation determined that the relationships between the prediction testing phase participants' self-rated levels of horse experience and the horse experience scores were statistically significant, strong, and positively associated for the laypeople group ($\tau_b = .703, p < .001$), the EAMHP group ($\tau_b = .533, p < .001$), and the NEAMHP group ($\tau_b = .647, p < .001$).

^c Equine assisted mental health professionals

^d Non-equine assisted mental health professions

* $p < .05$. ** $p < .01$

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 4

Table 4

Average Total Assessment Accuracy Scores Across All Professional Groups

Profession	N	Mean	Standard Deviation
Layperson	94	5.77	3.51
EAMHP ^a	55	5.12	3.53
NEAMHP ^b	51	3.52	2.43

Note. Presented here are the means and standard deviations for the average total assessment accuracy scores for each of the three profession groups analyzed in the prediction testing phase. A statistically significant main effect of profession was found *Welch's F*(2, 118.767) = 10.682, $p < .001$, $\eta^2 = .074$ with an observed power of .951 (alpha = .05).

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 5

Table 5

Average Total Assessment Accuracy Scores Across All Self-Rated Horse Experience Levels

Self-Rated Level of Horse Experience	N	Mean	Standard Deviation
None	37	3.92	3.34
Low	50	4.41	2.63
Medium	42	4.93	3.45
High	71	5.9	3.32

Note. Presented here are the means and standard deviations for the average total assessment accuracy scores for each of the four self-rated levels of horse experience analyzed. A statistically significant main effect of self-rated horse experience was found $F(3, 196) = 3.832$, $p = .011$, $\eta^2 = .055$ with an observed power of .814 ($\alpha = 0.5$).

A post-hoc Games-Howell analysis indicated that both the participants who said they have no horse experience ($p = .024$ with a mean decrease of -1.97, 95% CI [-3.75, -0.20]) and those that said they have a low level of horse experience ($p = .035$ with a mean decrease of -1.49, 95% CI [-2.90, -0.07]) scored significantly lower than those participants who said they have a high level of horse experience.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 6

Table 6

Results of Multiple Regression Analysis Pertaining to Horse Experience Scores and Profession

Variable	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	95% <i>CI</i>
Scored Horse Experience	0.028	0.014	2.063	0.04	[0.001, 0.056]
Reference Group: EAMHP ^a					
Constant	3.852	0.753	5.118	<. 001	[2.367, 5.336]
Laypeople	1.001	0.574	1.743	.083	[-2.134, 0.132]
NEAMHP ^b	-0.622	0.785	-0.793	0.429	[-0.926, 2.171]
Reference Group: Laypeople					
Constant	4.853	0.555	8.738	< .001	[3.757, 5.948]
NEAMHP	-1.623	0.638	-2.545	.012	[-2.881, -0.366]

Note. *N* = 200 participants (consisting of 94 laypersons, 55 EAMHPs, and 51 NEAMHPs). Profession was dummy-coded for this analysis. Profession and horse experience scores, as a set, statistically significantly account for variation in the total assessment accuracy scores, $F(3, 196) = 6.76, p < .001$. The R^2 for the overall model was 9.4% with an adjusted R^2 of 8%.

Repetitious comparisons have been omitted.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

ACCURACY OF HORSE AFFECT ASSESSMENTS

Table 7

Table 7

Average Total Assessment Accuracy Scores Across All Professional Groups including Experts

Profession	<i>N</i>	Median	<i>M</i>	<i>SEM</i>	<i>SD</i>	Minimum	Maximum
Layperson	94	5.63	5.77	0.36	3.51	-1.34	13.3
EAMHP ^a	55	4.95	5.12	0.47	3.47	-1.71	13.31
NEAMHP ^b	51	3.37	3.52	0.34	2.43	-1.19	8.7
Experts	12	17.02	16.59	0.73	2.52	12.03 ^c	19.98

Note. Presented here are the medians, means, standard error of means, standard deviations, and minimum and maximum scores for the total assessment accuracy scores for each of the three groups of participants from the prediction testing phase and the expert participants from the instrument development phase.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

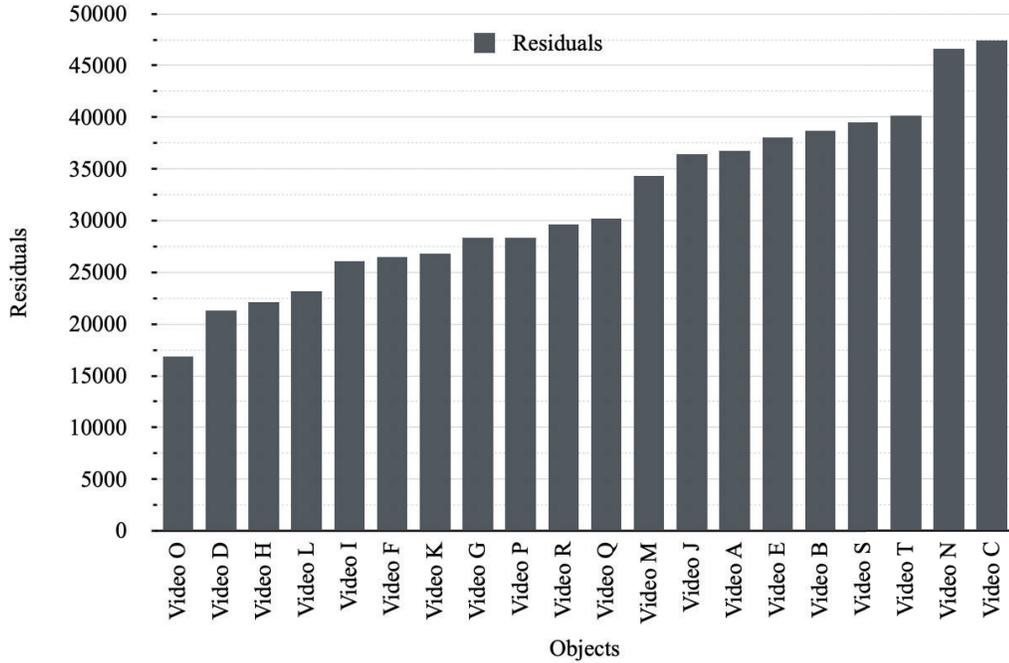
^c Because all of the terms generated by the expert participants in the instrument development phase of the study were used to construct the answer keys, against which the experts are being scored here in this exploratory analysis, this minimum score is not informative, as no expert was penalized for any of the terms they generated (and therefore they could never score negatively) unlike the three groups of participants from the prediction testing phase.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Figure 1

Figure 1

Instrument Development Phase: Generalized Procrustes Analysis: Videos Organized by Residuals



Note. This is a graph of the residuals by object (videos) found in the instrument development phase using a Generalized Procrustes Analysis with the videos presented in ascending order.

Appendix A

Overview of the Types of Therapy Which Involve Horses

Horses are often incorporated into therapy programs because they are seen to be congruent, behaving in ways that are directly reflective of their mental state and interactions with a human (Lac & Walton, 2012; Waite & Bourke, 2013; see also Ringhofer & Yamamoto, 2017). Due to their social nature, horses display uncomplicated responses to interactions with others (Waite & Bourke, 2013). It is believed that the immediate and direct feedback a horse gives through its visible behavioural response provides valuable information to the client regarding their own thoughts and behaviours (Merkies et al., 2018; Waite & Bourke, 2013).

Types of Equine Assisted Therapy: Therapeutic Riding

As the name would suggest, this form of equine assisted therapy predominantly takes place on the back of the equine as opposed to on the ground (Bass et al., 2009; Bona, 2020). While therapeutic riding may be used for a number of therapeutic goals, such as addressing emotional, cognitive, social, or other life related problems, it also has a dominant advantage when it comes to the exploration of physical goals and overcoming their associated barriers (Bass et al., 2009; Bona, 2020). By riding a horse, individuals with severe physical disabilities (for instance, multiple sclerosis or cerebral palsy) can gain access to an entire world of locomotion and physicality they may otherwise never get the chance to experience (Bass et al., 2009; Bona, 2020). Not only can this be psychologically supportive, but the movements of the horse's pace are not dissimilar from our own, and thus riding a horse can also assist in fitness and physical development (Bass et al., 2009; Bona, 2020).

ACCURACY OF HORSE AFFECT ASSESSMENTS

Types of Equine Assisted Therapy: Equine Facilitated Psychotherapy

The distinguishing feature of equine facilitated psychotherapy specifically, as compared to the other forms of equine assisted mental therapy, is the requirement that the acting psychological professional be a registered psychotherapist (Bona, 2020; Nilson, 2004). In equine facilitated psychotherapy, the therapy session is primarily conducted by two experts: One who handles the animal and is a credentialed equine professional, and another who is the therapist (Bona, 2020; Nilson, 2004). While it may appear that these experts are charged with separate tasks, due to the nature of the therapy it is important that these individuals are appropriately cross trained so that they have expertise in one another's fields (Bona, 2020; Nilson, 2004). However, sometimes a single expert may be fully cross-trained as both a therapist and an equine professional and may choose to work only with the horse and client in a unit of three (as opposed to a unit of four; Bona, 2020; Nilson, 2004).

Types of Equine Assisted Therapy: Equine Facilitated Counselling

Equine facilitated counselling is not dissimilar to equine facilitated psychotherapy. It may also be conducted as either a team of three or four (depending on if the equine professional and therapist are the same person, or two separate experts) with the major difference being the credentials of the therapist (Bona, 2020). As the name would suggest, in equine facilitated counselling the registered psychotherapist is replaced by a qualified and certified counsellor (Bona, 2020). Otherwise, the two are very similar. Equine facilitated counselling aims to treat an individual's psychological problems whether they be personal or clinical, with the aid of an equine (Bona, 2020). Activities involving the horse are selected to address specific goals and development as part of the

ACCURACY OF HORSE AFFECT ASSESSMENTS

client's therapeutic treatment plan (Bona, 2020). Such activities tend to take place on the ground rather than in the saddle and can involve activities like grooming and leading the horse (Bona, 2020). Talk-therapy is also utilized to a great extent in equine facilitated counselling (Bona, 2020).

Types of Equine Assisted Therapy: Equine Experiential or Assisted Learning

The main distinction between equine experiential/assisted learning as compared to equine facilitated psychotherapy is that it may be conducted by a trained facilitator (who may or may not be trained in mental health) rather than a registered counsellor or psychotherapist. Further, equine experiential/assisted learning are always conducted with two professionals (the horse expert and the facilitator), a horse, and either groups of or individual clients. Both equine experiential and assisted learning involve treating the horse not as a tool but as a co-therapist whose purpose is to help facilitate client self-reflection via interacting with them (Bona, 2020).

Equine experiential/assisted learning uses the horse's behavioural responses to encourage clients to engage in deep, meaningful, and honest self-reflections regarding their own thoughts, behaviours, and emotions (Bona, 2020). For example, if the horse spooks when the client approaches, the individual may be invited to reflect on why that happened and may come to realize they themselves were feeling frustrated. This may lead to the generalization that they often approach others in confrontational, angry, or frustrated manners. In turn, the client may go on to test alternative approaches with the horse that can then be practiced beyond therapy. Thus, learning in this type of therapy is believed to occur through self-reflection, transference, generalization, and application (Bona, 2020).

Associations

EAGALA

The core values of the EAGALA program include: making a positive impact in both human and equine lives; maintaining quality, professionalism, and excellence across all programs; upholding ethical standards and principles; fostering a collaborative culture among its members; and continuously seeking improvement and innovation (EAGALA, n.d.). Founded in 1999, it is a non-profit organization and was one of the first to create a concrete model for equine assisted therapy (EAGALA, n.d.).

Like equine facilitated psychotherapy, the model involves a therapist working alongside an experienced horse person in a team of four, with the horse and client (EAGALA, n.d.). One professional is the equine specialist, while the other is the mental health professional; these two roles are never filled by the same person in this model (EAGALA, n.d.).

Unlike the previously mentioned forms of equine assisted therapy, the EAGALA model is an exclusively ground-based approach in which riding never occurs (EAGALA, n.d.). Instead, the client's therapeutic program is supported by specific ground-based experiences, and through the use of the animal as a metaphor (EAGALA, n.d.).

Finally, the EAGALA model adopts a solution-oriented approach, in which the client is believed to already contain the answers/solutions to their own problems, which can be discovered through exploration and problem-solving (EAGALA, n.d.). As of 2019, EAGALA had 500 accredited programs across 40 countries in which they had served 49,945 clients with the help of 4,764 horses.

ACCURACY OF HORSE AFFECT ASSESSMENTS

PATH Intl.

Founded in 1969, the PATH Intl. has 813 centres across Canada and the United States, with nearly 7,000 members and serves 53,300 clients with special needs—including 6,000 veterans—every year (PATH Intl., n.d.-a). PATH Intl. supports its members in the provision of therapeutic/adaptive horsemanship; equine-assisted learning; incorporating equines into counselling, occupational therapy, physical therapy, psychotherapy, and speech-language pathology; and equine-assisted services for veterans (PATH Intl., n.d.-a). While PATH Intl. has an application process specific to prospective members who are already registered therapists/counsellors, both members with and without mental health professional credentials can become PATH Intl. certified after completion of the required prerequisites.

PATH Intl. holds the core values of: “promoting diversity and opportunity in equine-assisted services,” “providing a culture of safety, understanding and ethical treatment of both the humans and the horses,” “connecting and partnering with those who share the PATH Intl. vision in a mutually beneficial manner,” “sharing valued knowledge with our constituents to facilitate their success,” “promoting quality in all undertakings,” “encouraging and supporting creativity, inquiry and cutting-edge research,” “ensuring that all business is based on ethical principles and conducted with transparency,” “enhancing the value and credibility of the industry,” and “promoting an awareness of body, mind and spirit in equine-assisted services” (PATH Intl., n.d.-a, PATH Intl. Core Values section).

PATH Intl. does not specify rules regarding the incorporation of horses into therapeutic practices beyond stating that “licensed medical or behavioral health

ACCURACY OF HORSE AFFECT ASSESSMENTS

professionals [should] work within their scope of practice” and “the therapist determines how equine interactions, movement or aspects of the equine environment may enhance the client’s individualized treatment plan with the goals of improving the client’s sensorimotor, cognitive, behavioral, emotional or psychological function for improved health and wellness” (PATH Intl., n.d.-b).

Appendix B

Equine Welfare Risk Contributions of Improper Husbandry, Training, and Knowledge

Improper husbandry, poor animal training skills, and lack of knowledge all have been found to contribute to the development of undesirable behaviours in horses (Horseman et al., 2016; Minero & Canali, 2009). Unfortunately, such undesirable behaviours not only negatively impact a horse's mental and physical wellbeing but also greatly increase the animal's risk of being slaughtered (Minero & Canali, 2009).

In an archival analysis of the current literature on equine welfare, Minero and Canali (2009) summarized areas of concern and their implications with regards to horses. For our purposes we will focus on two: (1) behavioural problems and, (2) training and riding.

Behavioural Problems: Stereotypies

First, Minero and Canali (2009) considered the case of *stereotypies*. Also known as vices, stereotypies represent repetitious, seemingly purposeless behaviours almost akin to human compulsions (McGreevy, 2004; Waters et al., 2010). They include such unwanted behaviours as: stall-kicking, head-tossing, pawing, wood-chewing, and licking (to name a select few, for a more thorough list see McGreevy, 2004). Such behaviours are considered undesirable to the point where horses with stereotypies may not be permitted to be housed at some barns for fear, they may teach others (Mcbride & Long, 2001). It is not known exactly what causes stereotypies, but it is known that the performance of them “enables stressed animals to cope with stress” (Minero & Canali, 2009, p. 224). Thus, it is

ACCURACY OF HORSE AFFECT ASSESSMENTS

likely that management factors play a fundamental role in stressing and frustrating horses, leading to these undesirable behaviours (Minero & Canali, 2009).

For example, despite their extremely social nature and innate desire for locomotion (for which we prize them), it is extremely popular to confine horses to individual stalls (Yarnell et al., 2015). Isolation (such as use of individual paddocks or 24-hour stalling) practices, are often employed with the intention of preventing environmental injury to the animal (Minero & Canali, 2009). However, this protection comes at the cost of reducing the horse's ability to interact with conspecifics, which unfortunately can be a contributing factor in the development of stereotypies (Waters et al., 2010) considered indicative of poor mental welfare (Jørgensen et al., 2011; Waters et al., 2010). Unfortunately, stereotypies are often incorrectly treated through increased isolation, or through physical preventions (rather than addressing the source of the behaviour), which leads to further stress and frustration or provides mere Band-Aid solutions to the problem at hand (Minero & Canali, 2009).

Behavioural Problems: Training and Riding

The process of training and riding a horse also poses a context in which equine welfare may be compromised (Minero & Canali, 2009). Horses may be put in mentally or physically harmful situations if the rider overestimates the horse's cognitive and reasoning abilities, for example when a rider says something akin to "he did that on purpose," or "she knows exactly what she's doing to frustrate me" (Minero & Canali, 2009). Unfortunately, such statements are commonplace in many equine contexts. By making these statements, the rider and/or caretaker sets the animal up for failure by

ACCURACY OF HORSE AFFECT ASSESSMENTS

expecting it to think or act in ways that are beyond its capabilities (Minero & Canali, 2009).

A second way in which training and riding can impact horse welfare occurs when inconsistent signals or punishments are used (Minero & Canali, 2009). Horses—like pigeons, rats, mice, dogs, cats, fish, and all manner of animals—are best taught by someone with a clear understanding of learning theory (Minero & Canali, 2009). Too often riders and trainers lack knowledge of learning theory, and thus attempt to train horses using inconsistent signals, harsh punishment, and non-contingent punishment (Minero & Canali, 2009). While the latter two present welfare problems in and of themselves, overall failure to train a horse will likely result in behavioural problems (Minero & Canali, 2009; Horseman et al., 2016)—the reason 66.4% of horses between the age of 2 and 7 are sent to slaughter (Minero & Canali, 2009).

Clearly, potential equine welfare problems span many different contexts and come from many different sources. As the study conducted by Minero and Canali (2009) was purely archival, we will now turn our attention to a more qualitative investigation of the welfare issues facing horses.

Welfare Problems Facing Horses

Despite the consensus among adult Canadian equine industry members' that horses are capable of experiencing “positive and negative emotions, particularly pain and fear,” (DuBois et al., 2018, p. 1) it was found that their opinions on equine welfare rarely reflected these beliefs (DuBois et al., 2018).

Canadian equine industry members agree that there are welfare issues facing horses in the industry, and that equines at auctions and living on feedlots are the most at

ACCURACY OF HORSE AFFECT ASSESSMENTS

risk (DuBois et al., 2018). Interestingly, members agreed that inappropriately trained horses and horses with unknowledgeable owners or caregivers are particularly at risk of poor welfare (DuBois et al., 2018). In fact, owner ignorance and lack of knowledge were believed to be the most frequent cause “of reduced horse welfare” (DuBois et al., 2018, p. 7).

These findings were also observed in a survey of equine stakeholders in Great Britain (Horseman et al., 2016). This study found the leading root causes of horse welfare problems were attributed to (1) caregivers lacking knowledge, (2) a lack of finances, and (3) poor advice seeking behaviour of owners/caregivers who listen to other unknowledgeable resources or take bad advice (Horseman et al., 2016). However, inappropriate use of the horse (e.g., asking it to do something it is physically incapable of doing), employing training methods that involve punishment, and pushing a horse to work harder than it should were also identified as likely to contribute to compromised horse welfare (Horseman et al., 2016). Of particular interest, the stakeholders believed that horse welfare was likely to be compromised when humans misunderstood the horse’s behaviour, especially when stress or pain went uncorrected (Horseman et al., 2016). Such factors can often lead to the development of unwanted behaviours (another context identified in which horse welfare may be compromised) and the selling or auctioning of the horse may lead to even greater welfare infringement issues (DuBois et al., 2018; Horseman et al., 2016). Thus, training and the accurate interpretation of animal behaviour plays a role in equine welfare (Gronqvist et al., 2017).

The Potential Effects of Training on Equine Expressions of Affect

As the interpretation of an animal's actions is important for avoiding undesirable behaviours, the potential effects of popular training methods to mask a horse's affective state (upon which interpretations of behaviour are informed) must be considered.

Training and working (i.e., being ridden) are critical elements of a horse's life. It is therefore appropriate to question the impacts that lifelong training and working might have on equines, from their personalities to their expressions of affect.

Work, Personality, and Expressions of Emotion

Horses do not differ as much in their overall emotional levels as they do in the ways they express their emotions (Hausberger et al., 2011). The expressions of emotions observed in horses included gazing, touching, snorting, pawing, rolling, tail-raising, vigilance (when a horse freezes, head held high with intently focused ears and eyes as if standing watch), and locomotion (i.e., gaits) such as slow walking, walking, cantering, and the trotting passage (an exaggerated form of the trot that is correlated with anxiousness; Hausberger et al., 2011).

A study by Hausberger et al. (2011) revealed that horses assigned to dressage and high school (a higher level of dressage) were more likely to exhibit anxious and excited behavioural patterns (i.e., snorting, tail raising, vigilance; Hausberger et al., 2011). Dressage and high school work demands horses that are extremely sensitive to training cues, react immediately to them, are physically strong, and accurate in their movements (Hausberger et al., 2011). In these disciplines the horse must be always "in-control," with their movements and gait highly restricted (Hausberger et al., 2011). These demands of the work may explain why horses recruited for dressage (and high school) work tended to

ACCURACY OF HORSE AFFECT ASSESSMENTS

demonstrate behavioural patterns and personalities indicative of high anxiety and excitability (Hausberger et al., 2011) when compared to horses assigned to jumper or voltige (also known as equestrian vaulting) work.

Voltige involves the performance of gymnastics and acrobatics atop a moving horse (Horses of the Sun, n.d.). What is meaningful to the purposes of this thesis is that horses with quiet, calm personalities were the ones assigned to voltige work, and thus might be most representative of the population likely to be employed for EAT. The voltige horses were described as the quietest (i.e., calm and predictable) and were more likely than all others in the study to express themselves via rolling and slow walking (Hausberger et al., 2011). They also showed little fear when faced with a novel object (Hausberger et al., 2011).

This research demonstrated that a relationship exists between the nature of work a horse is employed for, the types of horses that commonly fit these roles, and their expressions of behaviour and emotional state. The high levels of anxiety and excitement demonstrated by the dressage and high school horses could signal a need for a closer evaluation of potential welfare concerns related to these sports, and such studies have been conducted and can be found elsewhere.

Training Methods

As previously mentioned, training methods and consistency of cues may also contribute to the compromise of horse welfare (Hal et al., 2008; Hausberger et al., 2011; Horseman et al., 2016; McGreevy & McLean, 2009; Minero & Canali, 2009; Visser et al., 2009). Research suggests that when young horses are *started* (a modern term for *breaking* a horse) using training practices that emphasize natural horse behaviour, clear

ACCURACY OF HORSE AFFECT ASSESSMENTS

communication, and acknowledge the equine's needs their technical performance does not suffer and they are less stressed (Visser et al., 2009). In contrast, when started using conventional or traditional methods (involving negative reinforcement and punishment), young horses performed the same as the prior group but were more stressed (Visser et al., 2009). If the purpose of training is to teach the horse to perform and obey commands, then the similarities in technical performance revealed by Visser et al. (2009) would suggest that trainers should select sympathetic training practices over outdated conventional methods, as the former seems to be less stressful for the young horse (Visser et al., 2009). Choosing to train using conventional methods, considering these results, may be potentially unethical and a threat to the horse's welfare (Visser et al., 2009). Further, traditional training practices are believed to be one of the top four potential causes for reduced welfare by Canadian equine industry members (DuBois et al., 2018).

The Ethical Concern of Learned Helplessness

Training can also have an immense impact on a horse's willingness to express itself through behaviours that would allow for human interpretation to occur. While there is little scientific work in the area, the potential for learned helplessness in horses is not out of the question (Hal et al., 2008). *Learned helplessness* is a psychological condition which results from a critical loss of control over one's environment (Webster, 1994). Through repeated exposure to pain or aversive stimulus, which is inescapable, an animal may enter a state of learned helplessness, in which even when an escape from the aversive stimuli becomes available, the animal instead lays down and endures, simply ceasing all attempts to cope (McGreevy & McLean, 2009). The aim of most traditional training of horses involves attempts by the human to gain complete control over the horse

ACCURACY OF HORSE AFFECT ASSESSMENTS

(Hal et al., 2008). The further back one goes in the history of horse training, the more evident it becomes that horses have historically been, and in some instances continue to be, coerced into a state of learned helplessness (Hal et al., 2008). Take first of all the term *breaking* which refers to teaching a horse to accept being handled and ridden by a human (Hal et al., 2008). A horse is considered successfully *broken* when they forfeit all control and demonstrate thorough compliance with the human counterpart (Hal et al., 2008). While most modern training seems to attempt to utilize a more appropriate negative reinforcement approach, punishment is also common in dealing with biting, bite threats, kicking, and kick threats (McGreevy & McLean, 2009). Examples of punishment in modern horse training involve, but are not limited to:

- *Rapping*, an illegal practice which involves striking a horse's front legs as it jumps so that the animal believes it has hit the fence and will be more likely to tuck its legs more next time (and thereby reduce the chance of downing a pole and receiving a penalty when showing; McGreevy & McLean, 2009).
- Bringing a horse who has refused a jump to stand in front of the obstacle, often with violent handling of the reins and once there, whipping it relentlessly until it jumps (McGreevy & McLean, 2009).
- Western pleasure horses are subjected to the *yank and crank* technique in which they are commanded forwards by the spur, while simultaneously jerked back by the mouth—making the painful stimuli inescapable (Hal et al., 2008).
- The use of metal nosebands lined with teeth deeper than the skin of a horse's nose which causes inescapable, albeit fluctuating degrees of pain regardless of whether the horse's actions are desirable or not (sold by Frenería y Romanas López, n.d.).

ACCURACY OF HORSE AFFECT ASSESSMENTS

- The practice of *soring* (which is an illegal practice yet still prevalent in the industry) in which gaited horses are made to wear heavy chains on their feet, while their hooves are often baked in chemicals. This is done to cause the animal such immense pain that relief only comes as their hoof leaves the ground, resulting in an exaggerated, competition-desirable gait known as “The Big Lick” (The Humane Society of the United States, n.d.).

Further, in many disciplines horses who exhibit demeanours and behaviours indicative of learned helplessness (such as being “flat,” “submissive,” or “quiet” which are relabelled as being obedient, non-argumentative, and agreeable) are desired and preferred (Hal et al., 2008).

Appendix C

How the Equine Affect Assessment Accuracy Indicator was Used to Score

Participant Affect Assessment Accuracy

Scoring Correct Categories

The frequency of each category identified for a video in the instrument development phase (i.e., correct categories) was divided by the total frequency of (correct) categories identified for the same video and multiplied by 5. For example, in video A, all twelve participants from the instrument development phase produced six terms that correlated with Category 13, and there were 28 terms in total produced for video A (across all 34 categories) across all the participants. So to calculate the weighted value of Category 13 as a correct category for video A, the following formula was used: $(6/28)*5 = +1.0714$. Therefore, if a prediction testing phase participant generated a term(s) that was categorized in Category 13 for video A, they would receive +1.0714 points.

Weighted Points Were Awarded for Correct Categories

Weighted point values were selected over unweighted, or rank ordered point values in order to best represent the accuracy of participants, as well as honour the prevalence of the affective states of the horses in the videos as identified by the expert participants in the instrument development phase. Thus, a precise picture of each participant's equine knowledge and assessment accuracy was captured, where those who were extremely correct in their assessments saw this reflected in their scores, and those participants who were egregiously off-mark were penalized appropriately.

ACCURACY OF HORSE AFFECT ASSESSMENTS

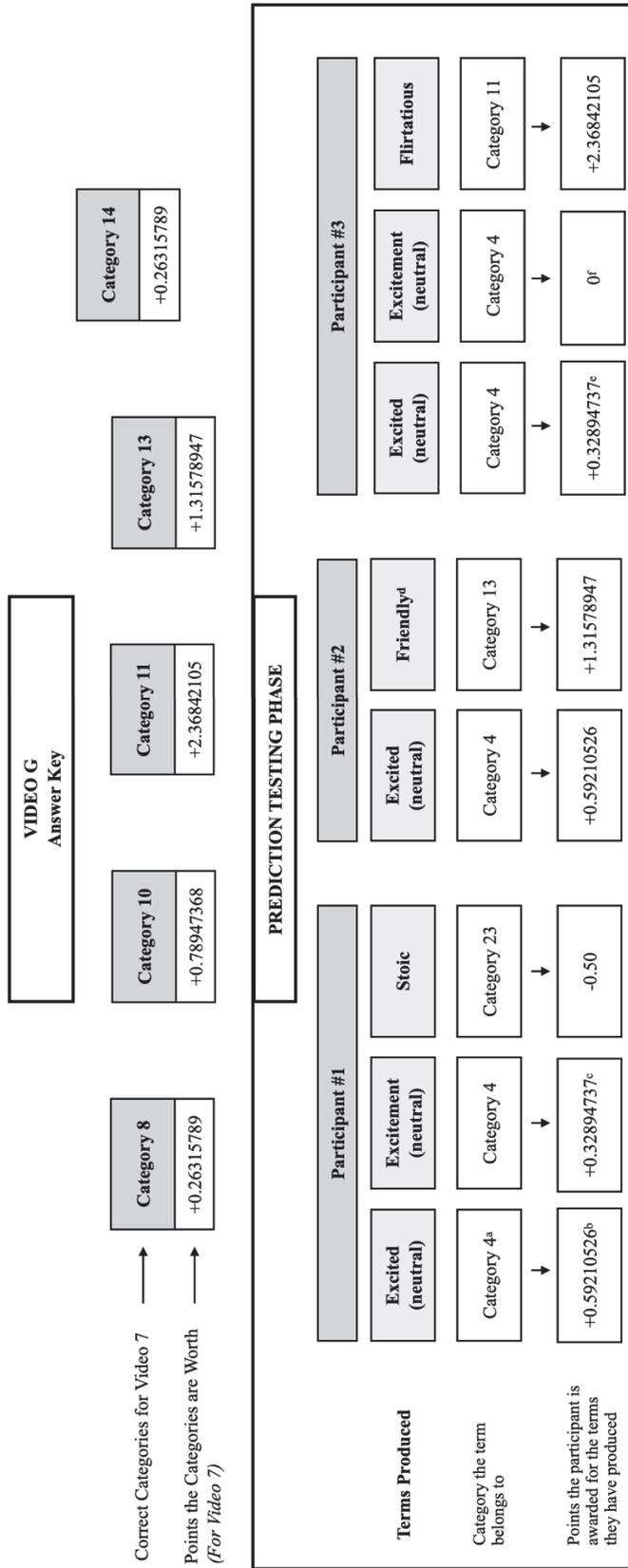
Regarding correct categories, associated point values were only awarded once per correct category identified, and not on a term-by-term basis. So, if in the above example the participant had provided two terms for video A that fit into Category 13, they still would have only received +1.0714 points not +2.1428 points.

Exceptions. Two exceptions to the point system were made for the terms “excited” and “excitement.” These terms belong to Category 4, Category 11, and Category 13 depending on the classifier they are assigned. If a participant produced the term “excited” or “excitement,” but due to interpretation of its description and the corresponding assignment of a classifier, it was assigned to a category considered incorrect for that video, the term was still awarded 1/4th of the point value(s) associated with the correct category/categories (for that video) which would have been identified if the “excitement” or “excited” terms had been given an alternate classifier. However, this bonus was not awarded if the participant produced other additional terms that belonged to the same correct category/categories as the correctly classified “excited” and/or “excitement” terms would have been.

For example (see also the Appendix C - Figure below), instrument development phase experts viewing video G identified Category 11 (via the term “excited [sexual]”) and Category 13 (though not due to an “excited” or “excitement” term) as correct categories. For video G, Category 11 is worth +2.368 points and Category 13 is worth +1.316 points. If a prediction testing phase participant watching video G did not produce the “excited (sexual)” term or any other term belonging to either Category 11 or Category 13 but did produce the “excited (neutral)” or “excitement (neutral)” term(s) (both of which fall under Category 4, a category worth no points in video G) the participant

Appendix C - Figure

Example Assignment of Bonus Points Associated with the "Excited" and "Excitement" Terms in the Prediction Testing Phase



Note. The terms "excited" and "excitement" belong to Category 4, Category 11, and Category 13 depending on the classifier they are assigned.

^a In the case of Video G, Category 4 is worth 0 points, however to account for potential classification errors of the term "Excited/Excitement," bonus points have been awarded once for Categories 11 and 13 to the participant.

^b Equal to the value of Category 11 divided by 4.

^c Equal to the value of Category 13 divided by 4.

^d Because the term "Friendly" belongs to Category 13, no additional bonus points are awarded to account for the "excited" term being classified into an Incorrect Category.

^e Equal to the value of Category 13 divided by 4.

^f Again, as the term "Flirtatious" belongs to Category 11, no additional bonus points are awarded for that category, even though the participant produced both the "excited (neutral)" and "excitement (neutral)" terms.

ACCURACY OF HORSE AFFECT ASSESSMENTS

would still receive +0.592 points for Category 11 and +0.329 points for Category 13. They would receive these amounts once for each category, regardless of whether they produced both “excitement (neutral)” and “excited (neutral)” or just one.

Had this participant additionally listed terms belonging to Category 11, like “flirtatious,” they would not have received the bonus for this category. The same is true if they had additionally produced terms that belonged to Category 13.

Rational for Exceptions. The decision to award these bonuses was made to account for potential inaccuracies in the process of hand-classifying the terms “excited” and “excitement.” As such, if an individual, via another term, was already awarded the points for the correct category that would have been associated with the “excited” and/or “excitement” terms—had they been classified differently—then there is no need to account for a participant being unjustly penalized, as they would have already earned the full points for the correct category.

Impact of Exceptions. It’s worth noting that this exception to the point system only applies to video I where “excited” was generated once in the instrument development phase and video O where “excitement” was generated once in the instrument development phase. The impact of these exceptions on the scoring of the prediction testing phase participants was minimal, impacting only 11 of the 210 total assessment accuracy scores (equal to 0.05%).

For video I, only one participant (a layperson, equal to 0.005% of the sample) earned additional points equal to +0.20 (to account for Category 13 and 4). For video O only seven participants (5 laypersons, 1 EAMHP, and 1 NEAMHP), equal to 0.03% of the sample, earned the additional 0.921 points (to account for both Category 11 and

ACCURACY OF HORSE AFFECT ASSESSMENTS

Category 13). Only one participant (a layperson, equal to 0.005% of the sample) earned the additional +0.592 points for Category 11; and two participants (1 layperson and 1 EAMHP, equal to 0.009% of the sample) earned the additional +0.329 points for Category 13 only.

For any other video, if a prediction testing phase participant generated the term “excited” or “excitement,” it would be clarified and categorized based on interpretation of its term description and scored based on the correctness of the associated category alone.

Why Exceptions Made for Only “Excited” and “Excitement.” Out of all the terms which have classifiers, only “excited” and “excitement” were selected for this scoring exception as the descriptions associated specifically with these two terms were in some cases particularly unhelpful in aiding in the classification of the terms. While the rules outlined in the Words Requiring Classification section of Appendix E were applicable in most cases, some participants used vague term descriptions such as “the horse is excited” and therefore interpretation and classification of the term as true to the participant’s intentions could not be guaranteed. These scoring exceptions were thus created to offset any potential errors introduced by the hand-classifying of such terms.

Scoring of Incorrect Terms

If a participant produced a term that fit into a category not identified for the video by the experts in the instrument development phase, that answer was considered incorrect and was awarded -0.5 points per occurrence. The value of -0.5 was derived from assigning a negative value for producing one incorrect term, and then taking the total number of incorrect terms a participant could produce on a video (which is equal to 10, as a participant can only ever produce up to 10 terms for a video) and dividing this by -5.0,

ACCURACY OF HORSE AFFECT ASSESSMENTS

so that the range of points a person could get on a video was balanced (i.e., -5.0 to +5.0). That is, $-5.0/10 = -0.5$. The decision to penalize each occurrence of an incorrect term, rather than one occurrence per category, was made to ensure the accuracy of the participants was appropriately reflected.

Rationale for Awarding Positive Points Per Category and Negative Points Per Term

Example Scenario Details

The following example scenario will be reused throughout this section to illustrate the rationale for the scoring system that has been selected for this study: Video X is worth 5 points, and the expert participants identified 5 categories to describe it. These categories are weighted as follows:

- Category 1: = +2.0 points
- Category 2: = +1.5 points
- Category 3: = +0.75 points
- Category 4: = +0.50 points
- Category 5: = +0.25 points

In these examples, participants 1, 2, and 3 will always produce the same answers, as this Appendix item is only concerned with the rationale for scoring. These participants will always produce the following categories for the above hypothetical video X:

- Participant 1: Produces 10 terms—one term that fits into Category 4 (correct) and nine incorrect terms.
- Participant 2: Produces three terms—one that fits into Category 5 (correct), one that fits into Category 4 (correct) and one that fits into Category 3 (correct), and no incorrect terms.

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Participant 3: Produces 10 terms—one that fits into Category 1 (correct), one that fits into Category 2 (correct) and eight incorrect terms.

Order of Accuracy. Before proceeding we must decide what the correct order among these participants is regarding their accuracy. Participant 1 is clearly the least accurate, as they have only produced 1/10 terms that are correct and the term they did identify correctly belongs to a category worth a low number of points.

Participant 2 produced fewer terms overall compared to Participant 1 and Participant 3, but because points are awarded based on correct categories and not term-by-term, this limitation of their vocabulary is less likely to impact their score negatively. Participant 2 produced 3/3 terms that were considered relevant to the video by expert participants in the instrument development phase and no incorrect terms. They have clearly identified the horse's affective state in the video correctly.

Participant 3 produced the same number of terms as Participant 1. Two of the terms they produced were considered highly relevant by the experts, and so they received high scores for these terms. However, 8/10 of the terms produced by Participant 3 were incorrect. While there is a possibility the participant felt a testing pressure to produce the maximum number of terms, of the 12 experts in the instrument development phase, not a single expert produced the maximum of 10 terms on any of the 20 videos. Further, of the 2000 sets of terms produced (200 participants in the prediction testing phase multiplied by the 10 videos) there was only one occurrence of a participant producing the maximum 10 terms for a video (and only 3 occurrences of participants listing 9 terms for a video).

If we assume that Participant 3 produced all 10 terms because they truly believed that they described the horse in the video, then while they did identify 2 highly relevant

ACCURACY OF HORSE AFFECT ASSESSMENTS

categories, they also produced 8 incorrect terms, and thus did not demonstrate a high level of understanding of the emotional state of the horse. Thus, the appropriate order of the participants, from worst to best accuracy, in this example should be: Participant 1, Participant 3, and Participant 2. This will be the assumption throughout the remainder of this Appendix item.

Example One Outcome: Zero Scoring for Incorrect Terms

First let us consider why a negative, rather than zero, score has been assigned for incorrect terms. In the absence of negative scores and the use of a 0-point system, the following situation could occur (also refer to Appendix C – Table 1):

- Participant 1: Produces 10 terms—one term that fits into Category 4 (+.50) and nine incorrect terms (no penalty). Their total accuracy score = 0.50/5.0.
- Participant 2: Produces three terms—one that fits into Category 5 (+0.25), one that fits into Category 4 (+0.50) and one that fits into Category 3 (+0.75), and no incorrect terms (no penalty). Their total accuracy score = 1.5/5.0.
- Participant 3: Produces 10 terms—one that fits into Category 1 (+2.0), one that fits into Category 2 (+1.5) and eight incorrect terms (no penalty). Their total accuracy score = 3.5/5.0.

In the situation below, there is no issue with Participant 1, as their score clearly reflects that they were less accurate than Participant 2 and Participant 3. But if we compare Participant 2 to Participant 3 there is a problem. As we have already discussed,

ACCURACY OF HORSE AFFECT ASSESSMENTS

Appendix C - Table 1

Example Scenario if Zero Points Were Awarded for Incorrect Categories/Terms

Video Being Assessed	Participant 1		Participant 2		Participant 3	
	Categories ^a	Score	Categories	Score	Categories	Score
1	Category 4	0.5	Category 5	0.25	Category 1	2
1	Category 6	0	Category 4	0.5	Category 2	1.5
1	Category 6	0	Category 3	0.75	Category 6	0
1	Category 6	0			Category 6	0
1	Category 6	0			Category 7	0
1	Category 7	0			Category 7	0
1	Category 7	0			Category 7	0
1	Category 7	0			Category 7	0
1	Category 7	0			Category 9	0
1	Category 7	0			Category 9	0
Total Accuracy Score		0.5		1.5		3.5

Note. In this particular example participants can only score between 0.0 and +5.0 on each video. In this example, the categories are weighted as follows: Category 1 = +2.0 points; Category 2 = +1.5 points; Category 3: +0.75 point; Category 4 = +0.50 points; Category 5 = +0.25 points.

^a Categories for which terms have been identified based on the participant's terms.

Participant 2 should have the highest accuracy score, followed by Participant 3, with Participant 1 scoring the lowest. In this scenario, the assignment of zero points for incorrect terms has resulted in Participant 3 being represented as much more accurate than Participant 2, which is incorrect.

Example Two Outcome: Negative Scoring for Incorrect Terms

If we take the exact same scenario but introduce the use of negative scoring for incorrect terms the following situation occurs (also refer to Appendix C - Table 2):

ACCURACY OF HORSE AFFECT ASSESSMENTS

- Participant 1: Produces 10 terms—one term that fits into Category 4 (+0.50) and nine incorrect terms (-4.5). Their total accuracy score = -4.0.
- Participant 2: Produces three terms—one that fits into Category 5 (+0.25), one that fits into Category 4 (+0.50) and one that fits into Category 3 (+0.75), and no incorrect terms. Their total accuracy score = +1.5.
- Participant 3: Produces 10 terms—one that fits into Category 1 (+2.0), one that fits into Category 2 (+1.5) and eight incorrect terms (-4.0). Their total accuracy score = -0.5.

Appendix C - Table 2

Example Scenario Where Negative Points Are Awarded for Incorrect Categories/Terms

Video Being Assessed	Participant 1		Participant 2		Participant 3	
	Categories ^a	Score	Categories	Score	Categories	Score
1	Category 4	0.5	Category 5	0.25	Category 1	2
1	Category 6	-0.5	Category 4	0.5	Category 2	1.5
1	Category 6	-0.5	Category 3	0.75	Category 6	-0.5
1	Category 6	-0.5			Category 6	-0.5
1	Category 6	-0.5			Category 7	-0.5
1	Category 7	-0.5			Category 7	-0.5
1	Category 7	-0.5			Category 7	-0.5
1	Category 7	-0.5			Category 7	-0.5
1	Category 7	-0.5			Category 9	-0.5
1	Category 7	-0.5			Category 9	-0.5
Total Accuracy Score		-4		1.5		-0.5

Note. In this example participants can score between -5.0 and +5.0 on each video. As before the categories are weighted as follows: Category 1 = +2.0 points; Category 2 = +1.5 points; Category 3: +0.75 point; Category 4 = +0.50 points; Category 5 = +0.25 points. Additionally, incorrect terms are awarded -0.5 points for every term produced.

^a Categories for which terms have been identified based on the participant's terms.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Here we can see that use of negative scoring for every incorrect term paints a much more accurate picture. Participant 1 had the lowest score because they produced many incorrect terms and only one, low-weighted correct category. Participant 2 scored the highest because they only produced correct terms and were therefore not penalized. Lastly, Participant 3 performed better than Participant 1 because they identified two high-weighted correct categories, and only 8 incorrect terms, but they still performed worse than Participant 2. This order is accurate to the scenario.

Example Three Outcome: Penalizing Per Category vs Per Term

Next let us consider why penalizing by term is more representative of participant accuracy than penalizing by category. The example scenario and participant terms remain the same. As the example two outcome above already presents a scenario in which penalization occurs term by term, a scenario in which participants are only penalized once per incorrect category will now be illustrated (see Appendix C - Table 3):

- Participant 1: Produces 10 terms—one term that fits into Category 4 (+0.50) and nine incorrect terms: four that belong to Category 6 (-0.5) and five that belong to Category 7 (-0.5). Their total accuracy score = -0.5.
- Participant 2: Produces three terms—one that fits into Category 5 (+0.25), one that fits into Category 4 (+0.50) and one that fits into Category 3 (+0.75), and no incorrect terms (no penalty). Their total accuracy score = +1.5.
- Participant 3: Produces 10 terms—one that fits into Category 1 (+2.0), one that fits into Category 2 (+1.5) and eight incorrect terms: two that belong

ACCURACY OF HORSE AFFECT ASSESSMENTS

to Category 6 (-0.5), four that belong to Category 7 (-0.5), and two that belong to Category 9 (-0.5). Their total accuracy score = +2.0

Appendix C - Table 3

Example Scenario Where Negative Points Are Awarded Per Incorrect Category Not Term Occurrences

Video Being Assessed	Participant 1		Participant 2		Participant 3	
	Categories ^a	Score	Categories	Score	Categories	Score
1	Category 4	0.5	Category 5	0.25	Category 1	2
1	Category 6	-0.5	Category 4	0.5	Category 2	1.5
1	Category 6	—	Category 3	0.75	Category 6	-0.5
1	Category 6	—			Category 6	—
1	Category 6	—			Category 7	-0.5
1	Category 7	-0.5			Category 7	—
1	Category 7	—			Category 7	—
1	Category 7	—			Category 7	—
1	Category 7	—			Category 9	-0.5
1	Category 7	—			Category 9	—
Total Accuracy Score		-0.5		1.5		2

Note. In this example participants can score between -5.0 and +5.0 on each video. As before the categories are weighted as follows: Category 1 = +2.0 points; Category 2 = +1.5 points; Category 3: +0.75 point; Category 4 = +0.50 points; Category 5 = +0.25 points. Additionally, incorrect terms are awarded -0.5 points only once per incorrect category produced.

^aCategories for which terms have been identified based on the participant's terms.

In this scenario, the order changes to (from highest to lowest score): Participant 3, Participant 2, and Participant 1. The scores are no longer reflective of the relative accuracy of these participants for these example scenarios.

Based on the above, it was determined that awarding incorrect terms with -0.50 points per term regardless of whether those terms belong to the same categories or not

ACCURACY OF HORSE AFFECT ASSESSMENTS

would produce results most reflective of participant accuracy. Participants were instructed to produce a minimum of one term to a maximum of ten terms for each video. These instructions, along with a review of the actual data indicates that participants understood the instructions and should not have felt pressured to reach for terms they did not feel were representative of the horse they were viewing. That is, participants should only have generated terms they truly believed were reflected in the video. If a participant selected an array of correct and incorrect terms this should be viewed as an indication of lack of understanding regarding the animal's affective states and associated with inaccuracy to the relevant degree.

Avoiding Penalizing Participants with Limited Vocabularies

If the participant produced 10 terms that all belonged to a single category, they received the point value for that category once, not 10 times. This decision was made to help prevent unfairly penalizing participants whose vocabulary only encompasses, for example, one word that fits into correct Category 1, when compared to a second participant, who for example, can list 10 words that fit into Category 1. Both participants are correct, and neither is more correct. So, Participant 1 should not be penalized based on the range of their vocabulary. Further, both participants in this case have correctly identified the affective state of the horse and in reality, would then be likely to respond appropriately, regardless of whether or not they would describe the horse as “happy,” “playful,” “joyful,” “blissful,” “friendly,” “bright,” “eager,” “confident,” “energized,” and “keen,” or just “joyful.” Only producing terms that belong to incorrect categories should be associated with repetitious penalties by term, as such inaccuracies are detrimental to the wellbeing and safety of the horse when manifested in the real world.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Appendix D

Categories and Terms

Appendix D

Categories and Terms Generated in the Instrument Development Phase and Prediction Testing Phase of the Study

#	Category Name	Instrument Development Phase Terms	Prediction Testing Phase Terms
1	Stressed / Concerned	Stress, Stressed, Concern, Concerned	Very Stressed, Slightly concerned
2	Insecure / Unsure / Nervous	Insecure, Unsure, Unsettled, Conflicted, Nervous, Uncomfortable (emotional), Uneasy	Nervous, Not sure, Uncertain, Uncertainty, Un-sure, Tense, Bracing, Comfort-seeking, Embarrassed*, Embarrassment*, Hesitant, Timid, Mildly Reactive, Nervous eating, Seeking assurance, Skittish, Shy, Rigid, Reluctant, Resistant, Unnerved, Retreating within self, Self-conscious, Managing*, Not confident, Humiliated, Tentative, Somewhat Ambivalent
3	Anxious/Worried	Anxiety, Anxious, Worried, Worry, Beckoning (seeking engagement in an anxious way), Distress, Distressed	Anxiety, Anxiety/worry, Anxious-worried, Anxious/worried, Worried/Anxious, Antsy, Anxious focused, Apprehension, Apprehensive, Devastated, Fidgety, Fretful, Troubled, On Edge, Perturbed
4	Alert / Attentive / Vigilant	Alert, Alert to activities in barn, Attentive, Aroused (non-sexual), Excited (alert, attentive), Excited (neutral), Excited (negative), Excitement (alert, attentive), Excitement (neutral), Excitement (negative), Anxious alert, Hyper vigilant, Vigilant, Watchful	Alert/vigilant, High alert, Alert and watchful, Alerted, Observant, On Alert, Hypervigilant
5	Agitated / Irritated / Frustrated	Agitation, Agitated, Irritated, Irritation, Agitation/Irritation/Frustration, Annoyed, Annoyance, Fussy, Frustrated, Frustration, Confined	Annoyed/irritated, Frustrated, Irritable, Bothered, Distraught, Exasperated, Flustered, Frantic, Grumpy, Insane, Insulted, Sulky, Miffed, Moody, Wound up, Worked up, Somewhat mad, Slightly irritated, Upset
6	Impatient / Restless	Impatience, Impatient, Restless, Impatience/Frustration	
7	Over-threshold	Over-stimulated, Overstimulated, Over stimulated, Emotional threshold, Over threshold	Disregulated, Dysregulated, Overwhelmed

ACCURACY OF HORSE AFFECT ASSESSMENTS

#	Category Name	Instrument Development Phase Terms	Prediction Testing Phase Terms
8	Curious / Interested / Investigative	Interested, Interest, Focused, Curious, Curiosity, Interactive, Investigating, Investigation, Searching, Inquisitive, Food-seeking	Curious, interested, Food seeking, Foraging, Inquisitive, Investigative, Slight curiosity, Explorative, Engaging with environment, Clever, Concentrated, Concentration, Attentive, Engrossed, Fixated, Focus, Inquiring, Intent, Intentional, Intrigued, Intrigue, Mesmerized, Mildly Curious, Wondering, Wondering - watching, Seeking connection (with environment), Responsive (to environment), Questioning, Seeking reward, Persistent, Engaged with task
9	Conspecific Focused Anxiety	Herdbound, Herd bound, Herd-bound, Mild separation anxiety, Separation anxiety, Socially Anxious	Anxious [separation], Separation-anxious, Social anxiety, Reactive
10	Social with Conspecifics	Social (towards conspecific) Horse social, Willing to interact (with conspecifics), Seeking connection (with conspecifics)	Engaging (with conspecific), Acknowledgement*, Buddy interested, Interested in other horse, Greeting other horse, Inviting (towards conspecific), Welcoming (towards conspecific), Loving (Towards conspecific), Belonging, Caring
11	Sexual Arousal and Related Excitement	Aroused (sexual), Flirtatious, Horny/In heat, Horny, In heat, In-heat, Estrous, Excited (sexual), Excitement (sexual), Expectant (sexual), Expecting (sexual)	Flirting, Anticipating (heat-based), Female in heat, Heat, In estrus, In season, Mare in heat, Mating seeking, Maybe in heat, Possibly in heat, Sexual desire, The mating, Activated, Attracted, Hormonal, Coquette*
12	Social with Humans	Engaged, Connected, Bonded, Social (towards human), Seeking connection (with human)	Engagement, Engaging (with humans), Affection, Affectionate, Appreciative, Attuned to person, Eager to connect (with human), Engaged seeking, Interaction with human, Invested, Inviting (towards human), Welcoming (towards human), Responsive (to human), Love towards human, Loving (towards human), Loved, Tender
13	Happy and Positive	Friendly, Happy, Bright, Eager, Excited (happy, positive), Excitement (happy, positive), Playful, Fun	Blissful, Exuberant - super happy, Happiness, Joy, Playfulness, Amused by object, Brave, Confident, Confidence, Cocky, Courageous, Proud, Elegant, Entertained, Energetic, Energized, Frisky, Good, Grateful, Gratitude,

ACCURACY OF HORSE AFFECT ASSESSMENTS

#	Category Name	Instrument Development Phase Terms	Prediction Testing Phase Terms
			Keen, Youthful, Spirited, Silly, Satisfied, Ready*, Lively, Pleased, Self-assured
14	Relaxed / Content	Content, Contentment, Relaxed, Relaxation, Relaxed foraging, Comfortable relaxed grazing, Normal rolling, Safe, Calm, At ease, Relief, Resting	Chill, Chilled, Comfort, Comforted, Contented, Ease, Increasingly comfortable, Laid-back, Lazy, Relax, Relaxed finally, Relaxed/at ease, Relaxing, Serene, Some relaxation, Very content, Very relaxed, Enjoy, Gentle, Kind, Trusting, Trust, Unafraid, Unbothered, Unconcerned, Unstressed, Watching - content, Unity, Quiet and comfortable, Reassurance, Secure, Secure relaxed, Settled, Soft, Leisurely, Mellow, Mindful, Safety, Self-soothing, Oblivious, Peaceful, Integrated
15	Surprised / Confused	Surprise, Alarm, Startled	Alarmed, Startle, Surprised, Confused, Confusion, Bewildered, Jumpy, Puzzled, Perplexed
16	Physical Enjoyment	Enjoyment, Pleasure, Physical pleasure (non-sexual)	Enjoying scratches, Enjoys head and ear grooming, Expressive*
17	Biological/Physical Sensations	Hungry, Itchy	Cold, Hunger, Hungry and cold, Tired, Sleepy
18	Physical Discomfort	Pain, Painful, Potential discomfort, Discomfort (physical), Physically uncomfortable, Unwell, Uncomfortable (physically), Sick, Respiratory sounds, Dull, Fidgety ill-at-ease	Ill, In pain, Possibly painful
19	Antisocial / Negative Interactions with Conspecifics	Anti-social (not wanting to be around others or engage in environment), Antisocial, Socially uncomfortable with, conspecific, Lacking socialization, Unfriendly with conspecific, Unfriendly	Fighting with conspecific, Jealous*
20	Protective	Protective, Guarding, Resource guarding, Resource guarding worry, Protective/Defensive, Defensive, Territorial*, Threatened*	Feeling threatened*, Protection
21	Dominance/Bossiness	Dominance, Assertive, Bossy, Asserting dominance, Leader, Leading, Dominating	Alpha, Dominant, Challenging, Posturing

ACCURACY OF HORSE AFFECT ASSESSMENTS

#	Category Name	Instrument Development Phase Terms	Prediction Testing Phase Terms
22	Aggression / Antagonistic Emotional State	Warning, Threatening, Aggressive, Aggressiveness, Aggression, Angry, Aggravated, Mad	Anger, Cranky, Judgemental, Ticked Off, Somewhat mad, Pissed off, Feisty, Confrontational
23	Stoic/Tolerant/Accepting	Stoic, Patient, Tolerant, Accepting	Obedient, Acceptance, Tolerance, Tolerating, Respectful, Submission, Submissive, Surrender, Resigned, Compliant, Docile
24	Thoughtful	Introspective, Thinking, Thoughtful, Quiet	Evaluative, Internalized, Processing, Pensive
25	Distracted	Distracted, Absent, Inattentive, Unaware	Oblivious, Occasionally distracted, Distraction, Occupied
26	Disinterested / Reserved	Bored, Boredom, Disinterested, Unmotivated	Not interested, Uninterested, Aloof, Reserved, Detached, Disconnected, Disconnected/uninterested, Distant, Flat, Independent, Indifferent, Neutral, Normal*, Self-engaged,
27	Positive Anticipation	Expectant (non-sexual), Expecting (non-sexual), Positive curious alert or anticipation, Positive anticipation, Anticipating (positive)	Anticipating fun, Anticipant, Anticipation, Calm but curious, Calm alert, Anticipatory, Expectation, Hopeful, In wait, Waiting,
28	Attention Seeking Behaviour	Attention seeking, Attention-seeking, Attention Seeker, Soliciting attention, Beckoning (seeking engagement in a calm way), Wants something	Seeking help, Seeking attention, Asking, Asking a question, Asking for love, Attention*, Trying to get attention, Wants attention, Needy
29	Lonely	Lonely, Longing, Withdrawn	Isolated, Isolated/separated, Yearning, Loss*, Abandoned
30	Unhappy	Unhappy, Displeasure, Not pleased, Discontent	Be unhappy, Depressed, Sad, Sadness, Disappointed, Grief, Sorrow
31	Wary, Guarded	Distrust, Leery, Wary, Suspicious, Skeptical, Guarded, Caution, Cautious alert	Distrustful, Distrusting, Wary/cautious, Suspicious and on guard, On Guard, Cautious, Alert - cautious, Mistrust, Mistrust of people, Non trusting, Not trusting, Untrusting
32	Fearful	Afraid, Fear, Mildly fearful, Spooky	Fearful, Frightened. Scared, Spooked. Unsafe, Vulnerable, Panic, Panicked, Lost
33	Avoidant	Avoidant, Avoidance, Fleeing (Conspecific or human)	Avoiding, Evasive, Declining/refusing, Rejecting connection, Invaded, Defining personal space, Wants to be left alone

ACCURACY OF HORSE AFFECT ASSESSMENTS

#	Category Name	Instrument Development Phase Terms	Prediction Testing Phase Terms
34	Seeking Access / Locational Discomfort	Seeking to leave the space, Wants out/access, Fleeing (space)	Confined, Claustrophobic, Anxious about space, Left behind, Trapped, Trapped/Separated, Cooped up
35	Not A Description of Emotion		Pregnant, Another 2 legger, Body signals, Listening, Listening and sensing, None, Whiney, Vocal, Talkative, Subtle, Tension release, Stress release, Smelling, Unnatural*

Note. Here the 35 categories and associated terms have been listed. Terms are organized by the phase in which they occurred (columns) and the category to which they belong (rows).

* Denote terms that, based on name alone may seem like they would belong to a different category, however, further clarification via the term descriptions provided by the participants supports their categorization as presented.

Appendix E

Instrument Development Phase: Scoring, Coding, and Categorization

Visual Analogue Scale Scores

For every term produced, the instrument development phase experts were instructed to rate the prevalence of that term on the associated VAS provided. These VAS scores are needed to conduct a Generalized Procrustes Analysis. For each of the 34 categories, for every expert participant, and for every video, a single VAS score was either calculated from a VAS score mean or, in instances where the participant generated no more than one term belonging to the relevant category (on a video-by-video basis), the original VAS score for the single term was used to represent its associated category. In instances where an expert generated multiple terms which belonged to the same category (ex: “happy,” “friendly,” and “bright” all belong to Category 13) for a single video, a mean VAS score was calculated and used.

If a participant forgot to provide a VAS score for a generated term, a mean score was calculated based on all other VAS scores provided by that same participant for the same video, regardless of the categorization of those additional terms. This new mean VAS score was then used to fill in the missing score. Only 24 VAS scores were Mean Substituted (equal to 4% of the 484 responses). Means calculated in this way were rounded up or down as appropriate to the nearest whole number.

Categorization of Terms

As a FCP model for term generation was selected, a great number of descriptive words could be produced by the participants. As such, categories needed to be created in order to avoid penalizing a prediction testing phase participant who may produce a

ACCURACY OF HORSE AFFECT ASSESSMENTS

similar but not identical word to an instrument development phase expert. By categorizing terms, it also becomes easier to comprehend the general affective themes of each of the videos, according to the expert participants.

To reduce human error, terms were categorized and grouped together based on their definitions and synonyms according to Merriam–Webster dictionary (n.d.-b) and through supplemental use of the Merriam–Webster thesaurus (n.d.-c), Gottman’s Feelings Wheel (Willcox, 2020), and Thesaurus.com (n.d.). In total, 34 categories evolved out of the experts’ terms from the instrument development phase and the contents of these categories can be seen in Appendix D.

Clarification via Term Descriptions

In several instances the term descriptions provided by the participants needed to be consulted to determine what was meant by the term. The term, with the aid of the term description, was then organized into the appropriate category using the Merriam–Webster dictionary (n.d.-b), and supplemental use of the Merriam–Webster thesaurus (n.d.-c), Gottman’s Feelings Wheel (Willcox, 2020), and Thesaurus.com (n.d.),

Words Requiring Classification

Several instances occurred, within and across the participants, where the same word (e.g., “excitement”) was used in different contexts (e.g., “alert/attentive excitement” vs “happy/positive excitement”). Part of the code developed was instructed to flag these terms by changing the cell’s colour, so that they were easily identifiable and could then be reviewed on a case-by-case basis.

All instances of repetition words (i.e., “aroused,” “beckoning,” “excited,” “excitement,” “expectant,” “expecting,” “fleeing,” “social,” “seeking connection,”

ACCURACY OF HORSE AFFECT ASSESSMENTS

“uncomfortable,” and “willing to interact”) were differentiated and categorized based on interpretation of the term descriptions provided in each case by the participants. Such words were then re-named to include a classifier (e.g., “excitement [happy, positive]).

The following list outlines the rules used to assign classifiers for each of these terms:

- The terms “expectant” and “expecting” were either classified as sexual or non-sexual in nature.
 - If the term description described the horse as being in heat, trying to interact with another horse, pheromones, squatting, and/or urinating it was classified as sexual.
 - The non-sexual classifier was assigned if the term description did not describe the horse as being in heat or aroused, etc.
- The term “beckoning” was either assigned the classifier of:
 - seeking engagement in an anxious way (which was only assigned if the term description described the horse as anxious, nervous, and/or uneasy), or
 - seeking engagement in a calm way (which was only assigned if the term description described the horse as relaxed, calm, gentle, or not in an anxious way).
- The terms “social,” “inviting,” “welcoming,” and “loving” were either classified as being directed towards humans or conspecifics.
 - The former was assigned if the term description mentioned a person, observer, or human.
 - The latter was assigned if the description mentioned another horse.

ACCURACY OF HORSE AFFECT ASSESSMENTS

- The term “uncomfortable” was either classified as physical or emotional.
 - The former was assigned if the term description mentioned pain, illness, and/or physical discomfort.
 - The latter was assigned if the term description lacked any mention of pain, illness, or physical discomfort.
- The term “fleeing” was either classified as relating to fleeing from humans and/or conspecifics or trying to flee from the space itself.
 - The former was assigned if the term description mentioned a human, horse, and/or other entity.
 - The latter was assigned if the term description described the horse attempting to leave, escape their environment, access another environment, and/or avoid, escape, or flee a physical space (such as a stall, pen, or paddock).
- The terms “seeking connection” and “engaging” were classified as either:
 - human focused (if the term description mentioned a person/human/observer),
 - conspecific focused (if the term description mentioned another horse), or
 - environmentally focused (if the term description mentioned the horse engaging with or seeking something from its environment, stall, pen, or paddock).
- The term “anticipating” was classified as either heat-based or positively associated.

ACCURACY OF HORSE AFFECT ASSESSMENTS

- The former was assigned if the term description described the horse being in heat, aroused, trying to interact with another horse, pheromones, squatting, and/or urinating.
- The latter was assigned if none of these elements were mentioned.
- The term “aroused” was classified as either sexual, non-sexual, or excited:
 - As with “anticipating” the sexual classifier was assigned if the term description described the horse being in heat, aroused, trying to interact with another horse, pheromones, squatting, and/or urinating.
 - Non-sexual was assigned if none of these elements were mentioned.
 - The excited classifier was assigned only if the term description described the horse as excited but made no mention of the horse being in heat, sexually aroused, trying to interact with another horse, pheromones, squatting, and/or urinating.
- The terms “excited” and “excitement” had five classifiers:
 - sexual (assigned if the term description described the horse being in heat, estrus, aroused, pheromones, squatting, and/or urinating),
 - alert/attentive (assigned if the term description described the horse as being stiff, tense, alert, watchful, focused, pacing, restless, autonomic arousal, impatient, investigating, eyes wide, hurried, head high, arched neck, squealing, pawing, whinnying, anticipation, energetic, looking, and/or sparked interest),

ACCURACY OF HORSE AFFECT ASSESSMENTS

- neutral (assigned if the term description made no clear indication that the horse was experiencing a state of excitement associated with sexual arousal or feelings of happiness/positivity),
- negative (was assigned if the term description described the horse as being, aggressive, agitated, anxious, upset, avoidant, tossing its head, blowing air and/or stressed),
- and happy/positive (assigned if the term description described the horse as happy, relaxed, lively, being eager, bright, ecstatic, merrily, cheerful, quivering lips, leaning in, engaged, romping, hopping, initiating touch [not described as sexual], enthusiastic, pleasurable [not described as sexual], prancing, joyful, and/or playful).

In both phases of the study, the other terms generated by the participant for the associated video were not taken into consideration when assigning classifiers unless the description associated with the term in question was so vague that classification was impossible without considering the other terms generated by the participant for that video. To honour the participant's assessment as accurately as possible, a term's associated term description was always checked first in isolation, and only if this description failed to allow the term to be classified were term descriptions from the other terms (produced for the same video by the same participant) taken into consideration.

If a term description explicitly indicated two classifiers were applicable (e.g., if a participant provided the term name "social" and the description "the horse appears to act in a social way towards the person and the other horse,") then a second term was

ACCURACY OF HORSE AFFECT ASSESSMENTS

manually added to that participant's answers. In this example, the participant's original term would receive the classifier "social (towards human)" and a second term, "social (towards conspecific)," would also be added to a previously empty term name field for that video. This was done to ensure a participant's full assessment was clearly represented in the scoring.

As a participant's ability to assess the nuances of the horse's emotional state is of core interest to this study, it was important that classifiers be used to parse emotions such as those listed above. If a participant only identifies a horse as, for example, "uncomfortable" and is not required to describe in what way the horse is uncomfortable (physically vs emotionally), then important information about the person's ability to assess the animal's emotional state is being lost.

Code Development

Once all of the terms from the instrument development phase had been categorized, a code was then created to relabel the terms generated by the experts according to their relevant categories. The result was that all of the terms produced by each expert, for each of the 20 videos, were organized into one of the 34 categories. While a code was perhaps not necessary to process the data produced by the small sample of expert participants in the instrument development phase, its development was crucial for handling the great amount of data produced by the prediction testing phase participants and helped to significantly reduce the chance of human error.

Appendix F

The Instrument Development Phase: Contents of the Demographic Form

Age						
(Presented in a drop-down menu)						
18–24	25–34	35–44	45–54	55–64	65–74	75+
Gender						
(Presented in a drop-down menu)						
Male	Female	Non-binary/other	Prefer not to say			
Profession Type						
Please provide a title or description of the work you do (example, horse behaviourist):						
[An empty field was provided]						
Please briefly describe your attitude or philosophy towards horse training in one to two sentences:						
[An empty field was provided]						

Appendix G

Instructions for use of the QBA Procedure and Free Choice Profiling Method

Step 1: The Video Clip

Watch the 1-minute video at the top of the screen when it is presented on the next page. You may rewatch this video if you wish at any time before advancing to the next video in this study. In this study, you will be presented with 20 videos, each 1 minute in length. It is vital you watch each clip entirely, at least once.

Step 2: The Qualitative Behavioural Assessment and Free-Choice-Profiling Method

After watching the video please think of 1-10 descriptive terms you feel best describes the expressive qualities of the horse in the video you just viewed.

Example: "happy"

Please only name one term per answer field. You will be provided with room to generate up to 10 terms. You may reuse terms you have generated in response to a previous video.

Step 3: Describe the Term

Next, you will be provided with space to explain the term, provide details about why you selected it, and what behaviours it describes. If reusing a term from a previous video, you may enter "see previous" or provide a new description.

Human Example: "Jim appeared happy because he was smiling and waving. Then he started laughing. Happy means to be in a positive state and feeling good."

Step 4: Visual Analogue Scale

Lastly, for each term, please use the slider to indicate the degree to which you felt your term was present in the video. The left end of the scale is labelled 'minimum' (this expressive quality is absent) and the right end of the scale is labelled 'maximum' (this quality could not be present more strongly).

Step 5: Advance to the Next Video

You are in complete control of how long you spend describing each video and can choose to move on to the next one by selecting "Next." You can return to a previous video and change your answers if you wish by clicking "Previous."

Reminder

After your data has been submitted, it cannot be withdrawn. You may withdraw your participation and discontinue the study at any time up until submission. To do so simply close your browser before you submit your responses, and they will not be included. If you choose to withdraw from the study, the responses you provided up to that point will be destroyed.

Appendix H

Visual Analogue Scale

In past studies that have used the QBA, the VAS has appeared in the form of a line between 100–125mm (depending on the study) which appears adjacent to the term that has been either provided (fixed list) or produced (FCP) by the participant (see Clarke et al., 2016; Fleming et al., 2013; Minero et al., 2009, 2016, 2018; Ringhofer & Yamamoto, 2017; Wemelsfelder et al., 2001). It is relevant that the QBA has been conducted in an online mode previously in the literature. Minero et al. (2016) conducted the focus group portion of their study using online delivery; however, it is unclear if the accompanying “Visual Analogue Scale of 125 mm length” (p. 149) was also delivered online or was provided in person during the focus group phase. In the same study, Minero et al. (2016) had participants during the farm visit portion use a “dedicated electronic Android application, specifically developed at SRUC for QBA automated data recording and analysis” (p. 149). To use the application participants “ticked the Visual Analogue Scale next to each descriptor [term] at the appropriate point” (Minero et al., 2016, p. 149).

Appendix I

Survey Design for the Instrument Development Phase and Prediction Testing Phase

Stimuli: The Videos

In the instrument development phase, 20 video clips (10 in the prediction testing phase) with an average runtime of 63 s (range of 54 s – 1 min 23 s) were prepared by editing footage of horse behaviour collected by myself. Only spontaneously occurring behaviours were recorded for the purposes of this study—no horse was ever coerced to obtain footage. The length of the videos was selected to reduce the participant time commitment to complete the study in order to increase participation and is aligned with video lengths used by Minero et al., (2016, 2018).

Each video clip presented footage of either an individual horse or several (wherein the participant was explicitly told to focus on a specific, easily observed animal). Within each clip the horse being observed would demonstrate a collection of behaviours that were considered similar (for example, a video in which a horse goes from a sleeping state to startled/alert state would not be shown, as these behaviours are dissimilar). In the instrument development phase, the 10 videos for which the expert participants most agreed on the affective states of the horses were then selected for presentation in the prediction testing phase.

Procedure

Unless otherwise specified, the procedure described below is identical for both phases. Any differences in method between phases have been clearly described and noted below.

Both Phases: Consent

First, participants were presented with a letter of consent as approved by the ethics committee. The letter varied slightly (regarding personal information that would be collected) depending on the phase. It was clearly stated that by clicking the “I Consent” button the participant was confirming that they: understood the form; were voluntarily consenting; had not yet commenced participation or previously participated; and had been given adequate time to consider the information and ask any questions they may have. It was also made clear that they could choose to withdraw their consent and discontinue the study at any time by closing their browser before submitting their responses. Due to the anonymous nature of the study, answers could not be withdrawn after they had been submitted.

Prediction Testing Phase Specific: After selecting “I consent” participants were asked if they currently live in Canada or the USA. Selecting “yes” would take them to the demographic form, selecting “no” would end the study as participation eligibility would not be met. Due to the nature of the sampling procedure in the instrument development phase this screening question was not necessary.

Both Phases: Demographic Form

Instrument Development Phase Specific: Next, participants were presented with a demographics form. In the instrument development phase of the study participants were asked to provide their age (18–24; 25–34; 35–44; 45–54; 55–64; 65–74; 74+), gender (male; female; non-binary / other; prefer not to say), and to self-describe their profession type in an open field. The instrument development phase experts were not asked to self-report their level of horse experience as this group was specifically selected based on

ACCURACY OF HORSE AFFECT ASSESSMENTS

their advanced horse experience and knowledge of horse behaviour. All participants in the instrument development phase were also asked to briefly describe their attitude or philosophy towards horse training in a provided open field.

Prediction Testing Phase Specific. In the prediction testing phase, the demographic form again gathered information regarding participant age and gender. These participants were asked further questions regarding their profession type and were presented with an additional horse experience questionnaire. The prediction testing phase demographic form and the horse experience questionnaire are presented in Appendix K and Appendix L respectively.

How Participants Were Assigned to The Profession Groups. First, prediction testing phase participants were asked the question: “Are you a psychologist, counsellor, and/or mental health professional?” If “no” was selected, the participant would be asked no further questions and was classified as belonging to the layperson group.

If “yes” was selected, then the participant was asked two follow-up questions. First, they were asked “are you a member of your provincial, state, or national regulatory/licensing/certifying mental health organization/agency/governing body?” (“yes” or “no”). If a participant answered “no” then their data was not included in the statistical analysis. This was to ensure the three groups of participants in the prediction testing phase were properly represented and to avoid potential issues regarding the inclusion of unlicensed, uncertified, and unregistered mental health professionals.

Second, participants were asked to select from a list of equine assisted related activities what statements (informed by Human-Equine Alliance for Learning, n.d.) applied to them. The number of “yes” statements presented here reflects an attempt to use

ACCURACY OF HORSE AFFECT ASSESSMENTS

the most inclusive language possible, so that an eligible EAMHP participant would not be misclassified as a NEAMHP on the basis of not agreeing with a more general statement such as “horses are involved in my practice.” Further, these statements serve to ensure eligible NEAMHPs are not misclassified as EAMHPs if the horse was not being employed to assist in a mental health therapeutic capacity. The list contained the following items:

- No.
- Yes, I work together with horses in my practice.
- Yes, I use a team approach in my practice where one team member is a horse.
- Yes, horses help facilitate corporate team building activities in my practice.
- Yes, horses help facilitate relationship counseling in my practice.
- Yes, I use a therapeutic riding model (AKA: Equine assisted activities, therapeutic driving, therapeutic vaulting) in my practice.
- Yes, I use equine-facilitated psychotherapy (AKA: Equine assisted psychotherapy, equine-facilitated mental health, equine assisted counselling) in my practice.
- Yes, I use equine-facilitated learning (AKA: Equine-facilitated experiential learning, equine-facilitated personal growth, equine assisted coaching, equine assisted personal development) in my practice.
- Yes, I use hippotherapy (AKA: Equine assisted physical therapy) in my practice.
- Other (with an open field).

If the items: “No,” “yes, I use hippotherapy,” or “other” (in which a non-equine assisted mental health activity was listed) were selected in isolation from any other items on the list, the individual was considered a NEAMHP. Selecting any of the other answers,

ACCURACY OF HORSE AFFECT ASSESSMENTS

either in isolation or not, resulted in a participant being identified as an equine-assisted mental health professional.

No participant selected “other” and left the associated field blank; had this occurred, the participant would have been designated as a NEAMHP in order to err on the side of caution. One participant only selected “other” and indicated they had previously conducted equine-facilitated psychotherapy—however, as the participant explicitly reported not having done so within the last year they were classified as a NEAMHP.

Prediction Testing Phase Specific: Horse Experience Questionnaire

Following the demographics form, participants were presented with a horse experience questionnaire with multiple questions aimed at assessing a participant’s level of horse experience. A detailed review of the questions included in the horse experience questionnaire as well as the relevant scoring of a participant’s horse experience are again thoroughly described in Appendix L.

The Instrument Development Phase: Instructions

After hitting “next,” participants were presented with a page with instructions for the use of the QBA process and how to use the FCP method (see Appendix G).

Participants were asked to read these instructions carefully and were also informed that the instructions would appear throughout the experiment. Participants also received instructions on how to mark the digital VAS for each term generated. The instructions informed the participants that they were free to list as few as one or as many as 10 terms for each video and that they would not run out of time because they controlled the pace of the study activity by clicking “next” when they were ready. In order to prepare participants for the length of the experiment the number of videos to be shown were also

ACCURACY OF HORSE AFFECT ASSESSMENTS

disclosed to the participant in the consent form (instrument development phase = 20; prediction testing phase = 10).

The instructions were adapted from other research in which the QBA process was used (Clarke et al., 2016; Fleming et al., 2013; Gronqvist et al., 2017; Minero et al., 2009, 2016, 2018; Wemelsfelder et al., 2000, 2001). Participants were instructed to watch the clip and then to “write down terms they felt suitably described the expressive qualities of the observed animals” (e.g., “happy” or “distressed;” Fleming et al., 2013, p. 5). Room was also provided for the participant to describe the generated term or provide further details about why they selected the term and what behaviour constituted it.

Prediction Testing Phase: Instructions

In the prediction testing phase this process was improved through the implementation of an additional instructional video (runtime of 3 minutes and 29 seconds) which accompanied the written instructions. This video introduced the participant to the layout of the study activity and reviewed the same information as included in the written instructions presented after the demographics form in both phases. The instructional video included a 13 second video clip wherein a person (with their permission) is presented watching a video on their phone and the term “happy” and the descriptions “this person appears happy because they were smiling and laughing” and “they were smiling and laughing” were used as examples. A timer was implemented so that the “Next” button only appeared after the duration of the instructional video had lapsed. This was to ensure participants had fully read/listened to the instructions before advancing to the study activity.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Both Phases: Draw

After the final video was presented and the participant had clicked the “Next” button, they were taken to a page that invited them to participate in an optional draw. If the participant wanted to enter the draw, they were instructed to click a link (which would open a one-page Qualtrics survey in a new window). Here they were told the odds for the draw (1 in 10 for the instrument development phase; and 1 in 15 for the prediction testing phase) and asked to provide their name along with a phone number and/or email to be contacted through if they won. They were also informed that they would need to complete a skill-testing question to claim their prize if they won and that their contact information would be kept confidential and would not be linked to their study data.

Appendix J

Sampling Procedure for the Prediction Testing Phase

Laypersons (achieved sample size $n = 94$; intended sample size $n = 50$) were recruited through self-selection using the University of Lethbridge's SONA system (approximately $n = 50$), the posting of recruitment ads to public and private Facebook Groups (Horse Human Bond; Equestrian Social Science [prev. Human-Horse Relations Research Group]), and a recruitment ad posted to my Facebook story, Instagram story, and TikTok account.

Registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices involve horses (achieved sample size $n = 55$; intended sample size $n = 50$) were also recruited through self-selection via the posting of recruitment ads to the public and private Facebook Groups: Trauma Informed Equine Assisted Services Networking Group; Equine Assisted Learning; Equine Assisted Psychotherapy and Learning; and, Equine Assisted Learning & Therapy in Atlantic Canada. Eligible participants were also identified through Google using the terms "equine-assisted" and "therapy" and/or "counseling," and a specific province or state. Personal and organization websites were then carefully reviewed and if eligibility was confirmed those individuals/organizations were contacted via either the email or phone number listed publicly on the website. Of the sample approached ($n = 386$), 14% completed participation in the study.

Registered, licensed, and/or certified psychologists, counsellors, and mental health professionals whose practices don't involve horses (achieved sample size $n = 51$; intended sample size $n = 50$) were recruited through self-selection via the posting of a

ACCURACY OF HORSE AFFECT ASSESSMENTS

recruitment ad to the private Facebook Group: Psychologists in the United States & Canada. Eligible participants were primarily identified through PsychologyToday.com as well as through public provincial psychologist directories (the Saskatchewan College of Psychologists; and the Psychological Association of Manitoba) and Google, using the terms “therapy” and/or “counseling” and a specific province or state. Personal and organization websites were then carefully reviewed and if eligibility was confirmed those individuals/organizations were contacted via either the email or phone number listed publicly on the website. Of the sample approached ($n = 745$), 6% completed participation in the study. All recruitment emails, posts, and other correspondents were approved by the ethics committee.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Appendix K

The Prediction Testing Phase: Contents of the Demographic Form

Age						
(Presented in a drop-down menu)						
18–24	25–34	35–44	45–54	55–64	65–74	75+
Gender						
(Presented in a drop-down menu)						
Male	Female	Non-binary/other	Prefer not to say			
Are you a psychologist, counsellor, and/or mental health professional?						
Yes			No			
(If they selected yes, this question would appear next)						
Are you a member of your provincial, state, or national regulatory/licensing/certifying mental health organization/agency/governing body?						
Yes			No			
Are horses involved in your practice? (Select all that apply)						
No.						
Yes, I work together with horses in my practice.						
Yes, horses help facilitate corporate team building activities in my practice.						
Yes, horses help facilitate relationship counseling in my practice.						
Yes, I use a therapeutic riding model (AKA: Equine assisted activities, therapeutic driving, therapeutic vaulting) in my practice.						

ACCURACY OF HORSE AFFECT ASSESSMENTS

Yes, I use equine-facilitated psychotherapy (AKA: Equine assisted psychotherapy, equine-facilitated mental health, equine assisted counselling) in my practice.	
Yes, I use equine-facilitated learning (AKA: Equine-facilitated experiential learning, equine-facilitated personal growth, equine assisted coaching, equine assisted personal development) in my practice.	
Yes, I use hippotherapy (AKA: Equine assisted physical therapy) in my practice.	
Other:	[with a provided open field]

Appendix L

Prediction Testing Phase: Scoring, Coding, and Categorization

Horse Experience Questionnaire

In the prediction testing phase, following the demographics form participants were presented with a horse experience questionnaire comprised of the following questions and answer options:

1. I have interacted with a horse as part of a horse-themed/focused program or structured activity.
 - a. Yes
 - b. No
2. I participated in that program or structured activity when I was a... (select all that apply)
 - a. Child
 - b. Teenager
 - c. Adult
3. I have participated in... (select all that apply)
 - a. Team building exercise(s)
 - i. "Once," "more than once," or "never"
 - b. Trail-riding experience
 - i. "Once," "more than once," or "never"
 - c. Summer camp
 - i. "Once," "more than once," or "never"
 - d. 4H Horse program

ACCURACY OF HORSE AFFECT ASSESSMENTS

- i. “Once,” “more than once,” or “never”
 - e. Other – text entry
 - i. “Once,” “more than once,” or “never”
- 4. How often did/do you interact with horses as a... (please select the answer that best describes you)
 - a. Child
 - i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,” “weekly,” or “daily”
 - b. Teenager
 - i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,” “weekly,” or “daily”
 - c. Adult
 - i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,” “weekly,” or “daily”
- 5. How often did/do you take horse-related lessons and/or training (please select the answer that best describes you)
 - a. Child
 - i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,” “weekly,” or “daily”
 - b. Teenager
 - i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,” “weekly,” or “daily”
 - c. Adult

ACCURACY OF HORSE AFFECT ASSESSMENTS

- i. “Never,” “once,” “less than once a year,” “yearly,” “monthly,”
“weekly,” or “daily”

6. Select all of the following that apply: I..

- a. “Have read books about horse training,”
- b. “Have groomed a horse,”
- c. “Have tacked up a horse,”
- d. “Used to and/or currently own my own horse(s),”
- e. “Used to and/or currently lease a horse,”
- f. “Have taken one or more courses about horse anatomy,”
- g. “Have taken one or more courses about horse behaviour,”
- h. “Grew up around horses,”
- i. “Have attended one or more horse clinics as a spectator,”
- j. “Have attended one or more horse clinics as a participant,”
- k. “Have taken college/university courses about horses,”
- l. “Have an equine-related job/profession/occupation,”
- m. “Have started/broken/gentled a horse (of any age),”
- n. “Have trained a horse (of any age),”
- o. “Used to compete in horse shows or competitions,”
- p. “Currently compete in horse shows or competitions,”
- q. “Used to interact with horses recreationally (outside of shows and competitions),”
- r. “Currently interact with horses recreationally (outside of shows and competitions),”

ACCURACY OF HORSE AFFECT ASSESSMENTS

s. “None of these options apply to me.”

7. I have worked and/or volunteered in a horse stable/barn as a...

a. “child,” “teenager,” and/or “adult,” or “never.”

Scoring

Participants received a designated number of points depending on the answers they selected to the questions above. For question 1, selecting “yes” was equal to 1 point, while selecting “no” was equal to 0 points.

For question 2, each answer was worth 1 point, so if a participant selected all three answers (“child,” “teenager,” and “adult”) they would receive 3 points. For question 3, answering “never” was equal to 0 points, “once” was equal to 1 point, and “more than once” was equal to 2 points per activity. So if a participant selected “team building exercise(s) - more than once,” “trail-riding experience - more than once,” “summer camp - more than once,” “4H horse program - more than once,” and “other - more than once” they would receive 10 points.

For questions 4 and 5, answering “never” was equal to 0 points, “once” was equal to 1 point, “less than once a year” was equal to 2 points, “yearly” was equal to 3 points, “monthly” was equal to 4 points, “weekly” was equal to 5 points, and “daily” was equal to 6 points per age category. So, a participant who answered: “child–daily,” “teenager–daily,” and “adult–daily” for both questions 4 and 5 would receive 36 points (18 per question).

For question 6, the answer “none of these options apply to me” was worth 0 points and selecting it would make the remaining answers un-selectable to help participants avoid accidentally selecting it. The remaining 18 options were each worth one point.

ACCURACY OF HORSE AFFECT ASSESSMENTS

For question 7, the “never” answer was equal to 0 points and the remaining three answers, “child,” “teenager,” and “adult” were worth 1 point. Therefore, a total of 71 points could be earned on the horse experience questionnaire.

Participants were also asked to self-rate their level of horse experience and were able to select from the following options: “no horse experience,” “low level of horse experience,” “medium level of horse experience,” or “high level of horse experience.” The participants’ self-rated level of horse experience was not used in the calculation of their horse experience score. Instead, a correlation was used to compare these self-rated levels of horse experience against their horse experience scores (which were calculated based on their answers to questions one through seven described above).

Term Processing

In the prediction testing phase the code developed in the instrument development phase was used once again to (1) categorize all pre-existing terms produced by the prediction testing phase participants, (2) to flag terms requiring classifiers, and lastly (3) to flag new terms that had not been previously produced in the instrument development phase (and were therefore uncategorized at this point). Regarding the latter, these flagged new words were first scanned to determine those for which punctuation marks, extraneous spacing, or misspelling had resulted in them being marked as new. Such terms (for example “interested.”) were then sorted into the appropriate categories (e.g., “interested.” would be sorted into the same category as “interested”). In instances of misspelling, the misspelt term was not categorized, but instead was corrected directly.

Categorizing New Terms

Next, new terms that had not yet been produced and categorized in the instrument development phase were reviewed by the author and sorted into the appropriate categories following the exact process described in Appendix E. For example, “joy” is defined as “a state of happiness or felicity: Bliss” (Merriam-Webster, 2023b) and thus was categorized alongside the word “happy” in Category 13 (see Appendix D). Any terms flagged by the code as needing a classifier were reviewed as described for the instrumental phase (see Appendix E, Words Requiring Classification section).

New Terms Requiring Clarification via Term Description

The remaining new terms for which further clarification was required before categorization could occur (e.g., “activated”) were then reviewed individually by the author on a case-by-case basis to determine to which category each instance of a new term belonged. This process was again identical to that described for the instrumental phase (see Appendix E, Categorization of Terms section), where the term’s description was also used to help clarify what the participant meant. In some instances, the new terms were added to the appropriate categories, whereas when a new term name had the potential to be confusing without the accompanying term description or if the term was very similar to an existing term, the new term name the participant provided was directly changed to an existing term name with the same meaning.

Handling Compounded Terms and Multiple Terms in a Cell

In the prediction testing phase, some participants listed multiple terms within the same term name field. In such situations, the following procedure was used to determine categorization on a case-by-case basis: First, if all of the terms in the combination (e.g.,

ACCURACY OF HORSE AFFECT ASSESSMENTS

“friendly–happy”) already existed within the same category but as separate terms (e.g., Category 13 contains both the terms “friendly” and “happy”) the combined term was placed into that category.

When At least One Term in the Combination Has Not Been Previously Categorized

If the combination contained one or more terms not previously categorized but the uncategorized term(s) fit into the same category as the already categorized term(s), then, after confirming with the Merriam–Webster dictionary (n.d.-b), the Merriam–Webster thesaurus (n.d.-c), Gottman’s Feelings Wheel (Willcox, 2020), and Thesaurus.com (n.d.), the terms would then be grouped in the same category. Take, for example, the terms “Expectant/Hopeful,” where “expectant” was not described as relating to sexual behaviour (and so belongs in Category 27), while “hopeful” (which was not previously categorized in the instrument development phase) is defined as “having or showing signs of eagerly awaiting something” and being synonymous with “anticipatory” (Merriam-Webster, 2023a). The term was therefore found to also fit into Category 27, alongside such terms as “positive anticipation”.

When Terms in the Combination Belong to Different Categories

If all of the terms in the combination (e.g., “irritated/defensive”) were already identified in the instrument development phase but belonged to different categories (e.g., “irritated” belongs to Category 5 and “defensive” belongs to Category 20), then the associated term’s description was reviewed on a case by case basis to: (1) determine if the participant was inarguably describing only one term (e.g., if the term description said: “The horse is very irritated”), in which case the term combination was categorized based on the primary term being described and relabelled as such; or (2) if it was either unclear

ACCURACY OF HORSE AFFECT ASSESSMENTS

or obvious that the participant intended to generate multiple terms to describe the horse, then the combined term was deconstructed and all terms were assigned to the appropriate categories. Take the example of “irritated/defensive,” the participant generated terms for both Category 5 and 20. So “irritated” would be entered into the field for term name 1 and “defensive” would be entered into the field for term name 2. Both terms would then be categorized by the code (as Category 5 and Category 20) and points would be assigned in accordance with the answer key for that video.

Rationale For Why All Terms in a Compounded Answer Were Counted

This decision to count multiple terms listed within the same field so long as the associated description was not inarguably defining a single descriptor of affect, was made in order to recognize the participant’s assessment. The value of honouring a participant’s assessment, which is, after all, the focus of this study, was determined to outweigh that of a participant not following the activity’s instructions correctly, and the discarding of answers on this basis would have represented a grave loss of data and resources.

Terms That Were Not Descriptions of Affect

Unique to the prediction testing phase of the study, some participants generated words that were not descriptors of emotion (e.g., “pregnant”). Unless the accompanying term description indicated an emotional state was in fact being identified (in which case the term was either relabelled with the term name from or simply added to the appropriate category), such words were assigned to a 35th category that was always worth -0.50 points. The addition of this category in the prediction testing phase had no impact on the findings of the instrument development phase, as all terms produced in the latter described emotions/affective states and were thus organized into categories 1–34.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Therefore, as no terms produced in the instrument development phase would have belonged to Category 35 had it existed, its creation in the prediction testing phase had no impact on scoring.

Scoring Accuracy of Equine Affect Assessments

It was decided that the points a participant could earn for each video would range from -5.0 to +5.0. Positive points were earned when a participant produced terms that correlated with the correct categories associated with that video (see Appendix C for more details regarding the weighting of points for correct categories). Negative points were earned when a participant produced terms that correlated with the incorrect categories for that video.

Appendix M

Free Choice Profiling vs Fixed Lists

In 2016, Clark et al. conducted a study to determine if observer ratings on the QBA were influenced based on whether the FCP or fixed list method was used by directly comparing the inter-observer reliability obtained for both. The results of the study indicated that observers scored the behaviour of 10 sows (female pigs) similarly, regardless of the method used. The degree of inter-observer reliability (i.e., consensus) obtained was similar for both groups, thus indicating that raters interpreted the observed behaviour analogously, using similar descriptive terms (Clark et al., 2016). Based on these findings, the choice between using FCP or a fixed list appears to come down to intention and the manner in which the QBA will be used. For the purposes of this study, inter-observer consensus, which can be assessed with either method, was the sole statistic of interest in the instrument development phase of this study.

Appendix N

**Instrument Development Phase: Video Selection Based on the Results of the
Generalized Procrustes Analysis**

After conducting the Generalized Procrustes Analysis, the categories and VAS scores produced by the 12 participants for each of the 20 videos were carefully hand-reviewed. This led to questions regarding the exclusion of videos J and S (as they were labelled in the instrument development phase; which were originally identified by the Generalized Procrustes Analysis as belonging to the group of 10 videos with the greatest residuals), and inclusion of videos D and G (as they were labelled in the instrument development phase; which were originally identified by the Generalized Procrustes Analysis as belonging in the top 10 videos with the lowest residuals).

The videos in question were then carefully reviewed based on knowledge I have of the context and horses in the clips (information unknown to the participants). The categories the participants had generated to describe these videos and the VAS scores assigned to those categories we also reviewed.

In both videos D and G some of the terms produced were questionable in their representation of the horses presented based on contextual knowledge of the situation and how the animal was behaving in the footage. On the other hand, the terms produced by the participants for videos J and S were not only accurate and highly representative of the horse's affective state as informed by contextual knowledge but were also in fact reflective of agreement among the expert participants.

It is possible that the quantity of categories identified in video J (e.g., "Category 1: Stressed/Concerned," "Category 2: Insecure/Unsure/Nervous," "Category 3:

ACCURACY OF HORSE AFFECT ASSESSMENTS

Anxious/Worried,” “Category 4: Alert/Attentive/Vigilant,” “Category 5: Agitated/Irritated/Frustrated,” “Category 7: Over-threshold,” “Category 9: Conspecific Focused Anxiety,” and “Category 31: Wary/Suspicious,”) and video S (e.g., “Category 5: Agitated/Irritated/Frustrated,” “Category 6: Impatient/Restless,” “Category 8: Curious/Interested,” “Category 13: Happy/Positive,” “Category 14: Relaxed/Content,” “Category 17: Biological/Physical Sensations,” and “Category 28: Attention Seeking Behaviour,”) led the Generalized Procrustes Analysis to consider them as having less inter-observer consensus. In reality all of these categories were representative of the horses’ discernable affective states and multiple categories were identified by the individual expert participants. As the terms produced for videos J and S were undoubtedly reflective of both consensus and the stimuli, whereas those terms produced for videos D and G were slightly questionable, it was decided that videos D and G would be replaced with videos J and S.

Based on the above findings, the final 10 videos selected for presentation in the prediction testing phase were as follows: videos F, H, I, J, K, L, O, P, R, and S. In turn, the following videos were excluded from the prediction testing phase because they either were identified by the Generalized Procrustes Analysis as having higher residuals (i.e., lower consensus) or because upon further inspection they did not reflect an adequate degree of consensus compared to those listed above: videos D, G, A, T, M, B, E, C, Q, and N.

The Interesting Case of Video Q

In this video a dapple-grey horse was laying down (upright, not on its side) in a stall. The horse appeared a bit sweaty, and its breathing was visible and somewhat

ACCURACY OF HORSE AFFECT ASSESSMENTS

accelerated. At the beginning of the clip the horse snorted once. Its ears were mostly still and off to the sides, and its lips twitched often as its head bobbed slightly. The context of this clip (unbeknownst to the participants) was that this horse had just been put away after being worked and soon after the video was taken the horse fell asleep.

Intriguingly, the participants generated appropriate terms to describe the horse all of which fell into either “Category 18: Physical Discomfort” or “Category 14: Relaxed/Content.” At first glance, the participants seem to be in agreement regarding the affective state of the horse in video Q, as only 2 out of 34 possible categories were identified. However, not a single participant generated terms for both categories. That is, participants either thought the horse was relaxing/content or they thought it was in physical discomfort, but none thought the horse was experiencing both. This is interesting because it reflects the usefulness of the Generalized Procrustes Analysis, which was able to recognize that despite the illusion of observer consensus, there was a split among the participants regarding which of the two categories described the horse’s affective state.

Appendix O

Preparing the Data for Statistical Analysis in the Prediction Testing Phase

To prepare the prediction testing phase data for analyses it was first extracted from Qualtrics. Next, all incomplete or potentially fraudulent participant data was flagged and then discarded. A participant's data was considered potentially fraudulent if the terms provided were: (a) not written in English (b) consisted of random letters and/or numbers but no real words, and/or (c) demonstrated word for word repetition to answers provided in other participant's data—which was then also disposed of. The data for which participants identified themselves as a as unregistered, unlicensed, and/or uncertified psychologists, counsellors, or mental health professionals were considered ineligible to participate and was also discarded (EAMHP, $n = 2$; NEMHP, $n = 6$).

Next, the data was analysed using the code developed in the instrument development phase. This resulted in (1) all terms previously generated by the expert participants being categorized, and (2) the flagging of terms requiring further analysis for classification and new words not previously generated in the instrument development phase. The flagged uncategorized terms were then processed, and the code was modified as described in Appendix L. The code was then re-run so that every term produced in the prediction testing phase data was appropriately categorized and relabelled. Based on the answers participants provided to the questions regarding profession, participants were also assigned to the appropriate group. Horse experience was also scored following the process described in Appendix L.

A second code was then run which checked every category identified by each individual participant for every video and awarded either the weighted point value for

ACCURACY OF HORSE AFFECT ASSESSMENTS

correct categories or a negative point value (-0.50) for incorrect categories as previously described (see Appendix C). The scores for each video were then tallied up for every participant. Participant professions, ages, genders, self-reported levels of horse experience, horse experience scores, assessment accuracy scores for each video, and total assessment accuracy scores were extracted and prepared for processing in IBM SPSS Statistics (version 29).

Appendix P**Normality of the Data Related to the Between Subjects One-Way Welch's ANOVA****Appendix P***Normality of the Total Assessment Accuracy Scores by Professional Group*

Profession	Skewness	<i>SE</i>	Kurtosis	<i>SE</i>
Layperson	0.217	0.249	-0.487	0.493
EAMHP ^a	0.269	0.322	-0.243	0.634
NEAMHP ^b	0.321	0.333	-0.374	0.656

Note. Presented here is the skewness, kurtosis, and associated standard errors found during assumption testing in preparation for running a one-way between subjects Welch's analysis of variance on profession. Total assessment accuracy scores were normally distributed across all three groups, as confirmed through the Shapiro-Wilk test ($p > .05$), and visual inspection of the histograms and Normal Q-Q Plots. Homogeneity of variance was violated, $F(2, 197) = 3.674, p = .027$.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

Appendix Q

Normality of the Data Related to the Between Subjects One-Way ANOVA

Appendix Q

Normality of the Total Assessment Accuracy Scores by Self-Rated Level of Horse Experience

Self-Rated Level of Horse Experience	Skewness	SE	Kurtosis	SE
None	0.562	0.388	-0.324	0.759
Low	0.363	0.337	-0.507	0.662
Medium	0.467	0.365	0.424	0.717
High	0.027	0.285	-0.371	0.563

Note. Presented here is the skewness, kurtosis, and associated standard errors found during assumption testing in preparation for running a one-way between subjects analysis of variance on self-rated level of horse experience. Total assessment accuracy scores were normally distributed across all four groups, as confirmed through the Shapiro-Wilk test ($p > .05$), and visual inspection of the histograms and Normal Q-Q Plots. There was homogeneity of variance, as assessed by Levene's test for equality of variances, based on the median, $F(3, 197) = 0.998$, $p = .395$.

Appendix R

Horse Experience Scores versus Self-rated Levels of Horse Experience

An examination of the average horse experience scores across the self-rated levels of horse experience grouped by the three professions (see Appendix R - Table 6) proves insightful for answering this question. In accordance with the findings of the correlation we indeed see that across all three groups horse experience scores increase alongside the self-rated levels of horse experience. We can also see that participants with high self-rated levels of horse experience from both the laypeople and EAMHP groups obtained similar mean horse experience scores.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Appendix R

Average Horse Experience Scores Across Self-Rated Levels of Horse Experience Grouped by Profession

Self-rated Level of Horse Experience	N	Median	Mean	Standard Deviation	Minimum	Maximum
Laypeople						
None	17	2	3.24	2.97	0	9
Low	23	16	17.96	9.01	2	40
Medium	20	44	42.95	14.35	19	65
High	34	48.5	50.12	11.12	31	69
EAMHP ^a						
None	-	-	-	-	-	-
Low	2	24	24	0	24	24
Medium	18	31	35.61	9.89	25	53
High	35	51	50.31	9.17	23	64
NEAMHP ^b						
None	20	4	4.05	3.66	0	17
Low	25	9	10.6	6.91	2	34
Medium	4	27.5	27.25	8.18	17	37
High	2	37	37	7.07	32	42

Note. Presented here are the medians, means, standard error of means, standard deviations, and minimum and maximum scores for the average horse experience scores of the prediction testing phase participants grouped by profession.

^a Equine assisted mental health professionals

^b Non-equine assisted mental health professions

However, the minimum and maximum horse experience score columns reveal a great degree of overlap between the levels within each profession. Further, the range of scores observed at the medium and high levels of the laypeople and EAMHP groups overlap substantially.

Laypeople who self-rated as having a low level of horse experience had a very extreme range of scores. These scores ranged from a minimum of 2 to a maximum of 40, with over half of this range also being observed at the medium level.

ACCURACY OF HORSE AFFECT ASSESSMENTS

Curiously, in the NEAMHP group, a sizable range of horse experience scores were observed at the self-rated level of having no horse experience (with a minimum score of 0 to a maximum score of 17). Additionally, in the laypeople and NEAMHP groups, those with a low level of horse experience obtained similar minimum horse experience scores to those associated with no experience.

Summary. The recurring theme of ranges within levels (particularly at the medium level) overlapping substantially with the ranges of other levels may help explain the lack of significant differences associated with the medium self-rated level of horse experience observed in the findings of the one-way ANOVA. In turn, the similar minimum scores of participants with no horse experience and a low level of horse experience (in the NEAMHP and laypeople groups) may help explain the lack of significant differences observed between these levels. Regarding the between-subjects one-way ANOVA—where participants who rated themselves as having no experience or a low level of experience were significantly less accurate than participants who rated themselves as having a high level of experience—there were no other meaningful differences in the assessment accuracy of the participants based on self-rated horse experience levels.

Decision. Given the current data, quantifying the range of horse experience scores associated with each level of horse experience is not possible. The overlapping of horse experience scores across levels has likely wiped out any potential effects other than those of the poles (which were observed). Further, because of this observed overlapping it seems likely that participants of truly different degrees of horse experience may have been incorrectly clumped together within the self-rated levels of horse experience. Thus,

ACCURACY OF HORSE AFFECT ASSESSMENTS

to accurately evaluate the effects of profession and horse experience on a person's ability to assess the affective states of a horse similarly to the experts, the horse experience scores were used instead of continuing with the self-rated levels of horse experience.