Comparison of Command-Based vs. Reality-Based Interaction in a Veterinary Medicine Training Application

Anton Franzluebbers[1]*, Simon Platt[2] and Kyle Johnsen[1]

[1] Virtual Experiences Laboratory, Department of Electrical and Computer Engineering, College of Engineering, University of Georgia, Athens, GA, United States, [2] Department of Small Animal Medicine & Surgery, College of Veterinary Medicine, University of Georgia, Athens, GA, United States

This paper presents a study evaluating alternative interaction styles for a novel virtual reality simulator proposed for veterinary neurology training. We compared a reality-based interaction metaphor, which is commonly used in virtual reality applications, to a command-based metaphor that reduced interactivity toward improving overall application usability. A cohort of 55 veterinary medicine students took part in the study, which took place at the veterinary school building. The study used a crossover design that allowed each participant to try both systems. Results suggested some correctable usability issues with the reality-based system, particularly the inclusion of haptic feedback for certain parts of the examination. A strong overall preference for the reality-based system was also observed. The study highlighted the potential of using both systems in tandem, with the command-based system being used prior to the reality-based system.

Keywords: virtual reality, simulation, training, reality-based interface, neurological examination

1. INTRODUCTION

One of the hallmarks of an immersive virtual reality (VR) application is the use of reality-based interaction metaphors (Jacob et al., 2008), which are also commonly called "natural" interaction metaphors. In a reality-based interaction system, the user performs intended actions according to how they are performed in the real world upon which the simulation is based. A common example of reality-based interaction is viewing a virtual environment within a tracked head-mounted-display (HMD). The user changes their viewpoint by ordinary head-motion. Reality-based object interaction is also quite common, facilitated in most cases by tracked hand-held controllers, which are followed by virtual counterparts (avatars). Squeezing a button on the controller while the avatar is near or pointing at an object is often used to grasp that object, causing the grasped object to follow the motion of the controller. Further reality-based metaphors exploit physics simulations that allow objects to respond, plausibly in many cases, to being released. One possible challenge within this paradigm, however, is when simulations fail, or have never been coded, to respond accurately to reality-based interactions. This may cause user frustration, or in training applications it could cause so-called "negative training transfer." This paper evaluates these possibilities relative to an alternative, command-based interaction scheme for a real-world application.

Our application was a physical examination of a virtual dog to be used by veterinary educators and their students (shown in **Figure 1**). As part of a physical examination, which is performed at intake and periodically while the animal is under observation, the user performs several procedures,
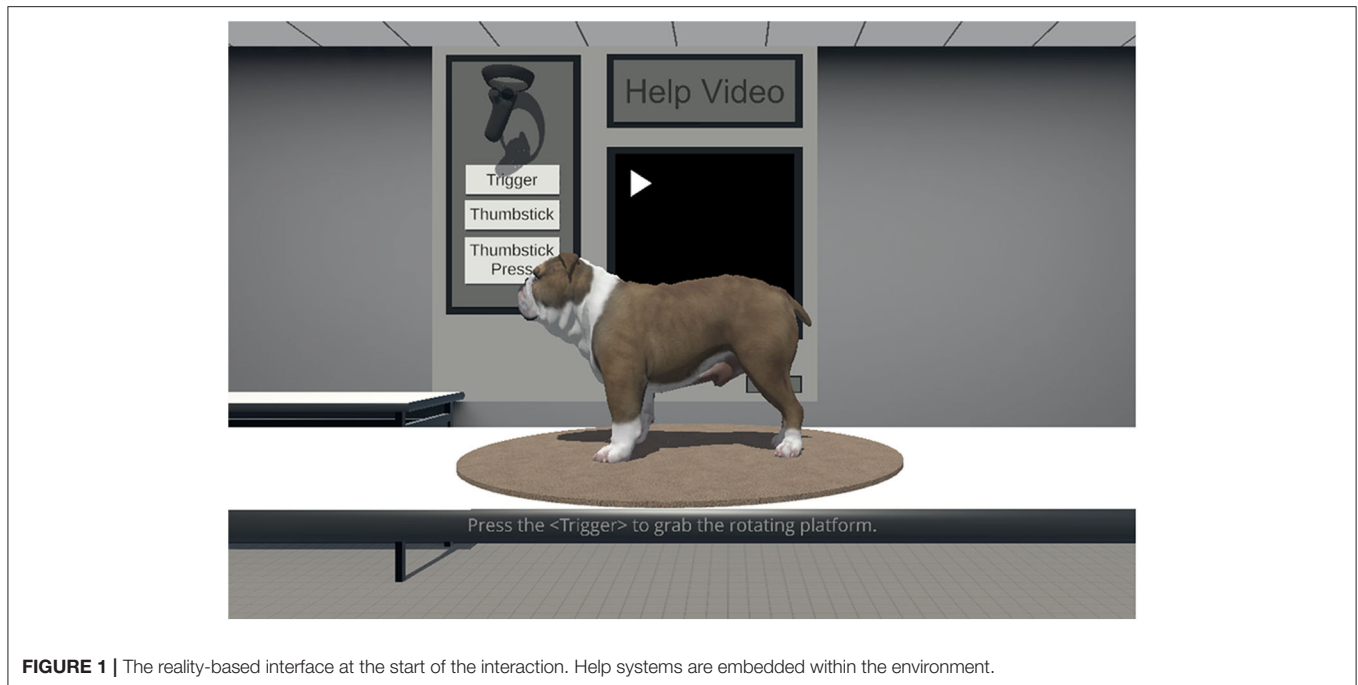
FIGURE 1 | The reality-based interface at the start of the interaction. Help systems are embedded within the environment.

each with a specific response from the virtual dog depending on its target simulation (e.g., normal or abnormal reflexes). Some responses have relatively straightforward behavioral simulations, such as the dog's eyes following a target. In this case, the simulation has a plausible response for nearly any input (tracked position) from the user, which is termed high representational fidelity in Whitelock et al. (1996). Other behaviors, however, are more complex, such as testing for the "hopping" response, which involves immobilizing one leg and having the dog hop on the contralateral leg. Correctly simulating this behavior through sensing and motor control is an active area of research (e.g., Zhang et al., 2018) and was deemed to be beyond the scope of this project due to resource constraints and the uncertain timeline for implementation. Instead, an interface and simulation approximation were used, which were only quasi-natural, and hence required user-interface training, making the application more difficult to deploy. Moreover, the inclusion of such a quasi-natural interface meant that there were actions the user may have tried to perform that did not produce a plausible response, such as trying to have the dog hop backwards instead of sideways. This was particularly concerning if the response or lack thereof was attributed to a physiological abnormality that was not intentionally simulated.

We explored an alternative, possibly complementary, interaction paradigm which restricted the user to a set of commands, enacted through a separate user interface (e.g., an embedded 2D graphical user interface). Commands caused actions to be performed by a third-party (in our case, a set of virtual hands). This approach had the advantage that the simulation input was known and thus the set of responses could be restricted to only plausible ones, even so far as to be pre-recorded animations. A significant disadvantage of this

approach, however, was that the user did not get to perform the examination, but instead watched the exam being performed. However, we considered the primary, most useful educational aspect of the system to be in recognizing abnormalities and understanding the examinations themselves. Training examination maneuvers was a secondary application goal, which could be facilitated by providing real, live dogs for practice (albeit without the abnormalities).

To evaluate the usability of the reality-based interaction system, we designed and conducted a user study where we asked veterinary students to perform neurological examinations using each interface, comparing their performance during the examinations and asking for their opinions about each interface. Our primary aim for this study was to determine if the considerable effort spent creating simulations that could respond to reality-based interactions was worth the investment, or if the less complex and expensive command-based system held similar value for the application, particularly given that the VR system would still, in theory, provide a sense of presence and stereoscopic viewing through visual immersion. In addition, a command-based system could practically be implemented without motion controllers, which add significant costs to a VR system: acquiring and maintaining the technology, the physical space required, and in ensuring the safety of immersed, moving VR users and others.

2. RELATED WORK

The term "reality-based interaction" was introduced in Jacob et al. (2007, 2008). It describes user behaviors with a computer interface that are based on corresponding, pre-existing, real-world behaviors. For example, the pinch to scale down

interaction metaphor may be tied to squishing something with one's fingers. Increasing the proximity of reality-based interaction metaphors has been a goal for immersive VR since its inception in Sutherland (1965), and though the "ultimate display" has yet to be constructed, visual and audio fidelity have increased dramatically (Brooks, 1999), and with this, perhaps, the emphasis on reality-based interaction metaphors. In other words, the more a virtual entity looks and sounds like a real-world entity, the more the initial, default behavior of a user with that virtual entity is likely to be based on how they would interact with the real-world counterpart, as seen in Reeves and Nass (1996). Further reality-based interaction is likely based on the behavioral realism of the entity in response to user interaction. For example, some studies have shown that it is important, at least for social constructs, such as co-presence, to match the appearance of an avatar to its behavioral realism, i.e., low-fidelity appearance with low-fidelity behavior and high-fidelity appearance with high-fidelity behavior (Bailenson et al., 2005; Vinayagamoorthy et al., 2005).

Reality-based interaction in VR is, in general, non-trivial to support, as seen in Stuerzlinger et al. (2006). Virtual objects do not, by default, behave realistically in response to input, which is also not, by default, realistically mapped to actions taken in the virtual environment. The developer must consider an array of factors that may influence how users expect objects to behave in response to input, such as the available computational resources, application goals, real-time requirements, input and output device-limitations, and development time. These challenges are also not unique to VR. Research and development efforts in the more general simulation field have perpetually struggled with this problem, such as in Scerbo and Dawson (2007). Moreover, there is mounting evidence, particularly in medical education, that the general preference for higher fidelity simulators in training and education is merely a product of "naïve realism" that is not backed by evidence. Increased simulation fidelity is directly tied to increased costs, and as such, decreased availability (Norman et al., 2012). Furthermore, there is a notable lack of scientific consensus that fidelity along any dimension is tied to improved outcomes (Norman et al., 2012). Indeed, some studies have shown that higher fidelity in early training may actually *inhibit* conceptual understanding (e.g., Schoenherr and Hamstra, 2017).

Our research lies predominantly in the medical education domain, which, like the aviation industry, has a long history of simulation-based training. Nearly all medical students receive some form of simulation-based training (e.g., Issenberg and Scalese, 2008; Motola et al., 2013). More recently, with the advent of relatively inexpensive, high quality VR systems, many other potential training opportunities are emerging that share similar characteristics and thus face difficult design considerations, specifically, applications that deal with the inspection, examination, and testing of physical entities where accurate, real-time simulations are not available or require considerable expense to implement. However, if a reality-based interaction metaphor is applied, users have a high degree of freedom of possible actions, which means that the virtual environment may be expected to respond to a near-infinite set of possibilities. For example, the purpose of many science and engineering education laboratory experiences is for students to gain experience measuring the real world and testing theories. These have notably been studied for over two decades in Dede et al. (1996) and Kaufmann et al. (2000) as possible VR applications, and are now commercially available (e.g., Labster [labster.com/vr]). Another example can be found in training for aircraft or vehicle maintenance tasks, where recognizing the need for maintenance may involve visual inspection and the use of various sensing devices that must be placed and operated appropriately to obtain accurate measurements (Vora et al., 2002; Vembar et al., 2005).

As described earlier, either from resource constraints or lack of knowledge, designers of the above applications must make approximations to the interface, simulation, or both when delivering a product. Our study, described in detail in the following sections, was designed to evaluate this tradeoff, where we systematically implemented a constrained, menu-driven interface relative to a reality-based interface used to perform a range of physical examination tasks. We evaluated differences in end-user performance, behavior, and attitudes between them. The results, while obtained within the veterinary education domain, have been generally informative of the costs, opportunities, and threats of supporting reality-based interaction, which we find applicable to a wide range of training and education scenarios.

3. MOTIVATION AND APPLICATION

Neurology is a standard subject for all veterinary students, which includes basic science and practical experience components. Veterinary students are learning the necessary anatomy, physiology, and examination tools and techniques, often in tandem, focusing on *normal* and later *abnormal* findings one might encounter in practice. Traditional training methods for the veterinary neurological examination of a small animal, such as a dog rely on a blend of classroom and laboratory materials and exercises. The process of an examination is described, shown through pictures and videos, and demonstrated and practiced on live animals. As the availability of animals with actual disorders may be limited, there have been efforts to provide virtual dogs that can portray abnormalities and respond to physical examinations. For example, Bogert et al. (2016) described a desktop-based VR simulator capable of responding to physical examinations and demonstrating a range of cranial nerve disorders, with a field study suggesting that the simulator could be effectively used as a part of knowledge assessments alongside text and video patient presentations. Similar efforts have also been recorded in the medical literature, such as Ermak et al. (2013), though a major difference between human and animal neurological examination lies in doctor-patient communication. The veterinarian must rely relatively more on physical signs and responses that can be elicited without requiring much cooperation and understanding on the part of the patient. These include stimulated reflexes and instinctual behaviors, such as walking and following objects.

With respect to previous work, our efforts represent higher immersiveness and functionality. These two aspects

are coupled, in that cranial nerve disorders primarily affect the face and eyes, which are presented well on an ordinary monitor without requiring 3D manipulation of the viewpoint or significant judgments of depth. This is particularly true of disorders affecting eye motion, which were the disorders presented during previous studies. In reality, a full neurological examination requires significant spatial movement to perform all examinations. Initial attempts at recreating this experience on a desktop monitor led to a rather complex interface to perform all maneuvers with the animal, including inconsistent use of interaction metaphors between examination components. Particularly, some used command-based metaphors for complex, two-handed maneuvers, while others used (constrained) reality-based metaphors, such as eye motion. Attempting to keep the reality-based metaphors consistent became increasingly difficult as the types of tests increased. In addition, we observed that providing a 3D viewing metaphor was challenging, and that when combined with a reality-based metaphor, the monoscopic 3D desktop view made certain depth judgments harder than in the real world, such as when using the reflex hammer. These issues motivated a switch to using a more-immersive, tracked, stereoscopic HMD with tracked joysticks as the user interface, and to the comparison of command-based vs. reality-based interaction metaphors for performing the examination. The HMD afforded an immersive viewpoint while the tracked controllers could be effectively used for either interaction metaphor. The additional degrees of freedom of the motion controllers over the mouse allowed a high degree of consistency in supporting the reality-based interaction between procedures, as described in the next section.

4. SYSTEM DESIGN

A virtual dog (an English Bulldog) was created that was the central focus of the neurological examination. The 3D model of the dog was purchased from a large online supplier, and the model was custom rigged and partially animated in Autodesk Maya and Unity3D to support full body and facial motions needed for the neurological examination. A custom simulation was made to drive the behaviors of the dog, which were parameterized to allow for abnormalities. Rather than focusing on the underlying disorders that could cause various physical signs, we parameterized the array of physical signs that may be observed and tested. This allowed the educator to create arbitrary virtual patients by specifying the physical signs, rather than just a few disorders that would exhibit particular signs. For example, we could encode a range of facial droop, reduction in control over a specific eye muscle, or limited ability to blink. An additional value of this approach is also that it allows ordinary variation with animals to be simulated, i.e., some differences are not abnormalities, such as slightly different sizes of pupils.

Users view the virtual dog with a tracked HMD, which for the present study was a Samsung Odyssey variety of the Windows Mixed Reality reference design. The Samsung Odyssey uses dual AMOLED displays, one for each eye, and each with a 1,440 × 1,600 pixel resolution at a 90 Hz refresh rate. For tracking, it uses a fusion of inertial sensors and an integrated, inside-looking-out optical head tracking system. This allows for a reality-based locomotion (real walking) and viewing metaphor. Furthermore, the HMD cameras track the position and orientation of light constellations on two bluetooth-connected hand-held joysticks, which are fused with further inertial tracking within each controller to provide quality, relatively low latency tracking of the joysticks. While exact performance characteristics are not available in the literature, we note that, for the present application, there was little difference in tracking performance between this system and an Oculus Rift or HTC Vive system, which were also possible to test through SteamVR. Notably, both the Oculus and HTC systems required the placement of external devices (cameras and emitters) for tracking, requiring more physical space and setup time. This is seen as an advantage by our veterinary education collaborators for simplicity. For similar reasons, we ensured that we could maintain a 90 frames per second rate throughout all tests and viewpoints using a gaming-laptop (NVidia Geforce 1070 GPU), making the system highly portable to different classroom environments.

The typical neurological examination consists of several parts, organized as tests to elicit reactions from the animal that are evaluated according to their degree of normalcy. The tests are performed in an examination room. As such, we designed the interaction to take place in a virtual examination room that is decorated similarly to real ones. At the start of the interaction, the virtual dog is standing on a turntable platform mounted on a table at the center of the room (see **Figure 1**). The user has the ability to adjust the height of the table and to spin the turntable as needed. The turntable reduced the amount of required walking and hence opportunities to hit things in the real environment or get tangled in the headset cable. The height adjustment allowed the user to set a comfortable head angle to view the dog. These supports would not be needed with more comfortable, untethered devices.

We implemented the most common tests that are performed, a list of which can be seen in **Figure 2** and are described below. For the purposes of the present study, we created two interaction modes for each part of the neurological exam, one that followed a command-based interaction metaphor (CMD mode) and one that followed a reality-based interaction metaphor (RBI mode). These differences primarily manifested themselves after selection of a particular test (e.g., patellar reflex testing). The various tests appeared on a menu tracked with one controller, and the user selected a test by using laser raycasting tracked from the other controller to intersect the button and pulling the controller trigger to activate the test (see **Figure 2**). In addition, controlling the table was consistent with each metaphor. In RBI mode, the height could be adjusted by "grabbing" the table and moving it up and down, while the CMD mode had dedicated buttons on the controller to raise and lower the table. Similarly, turntable rotation was performed by grabbing and rotating the turntable in the RBI mode, and performed by clicking on dedicated left and right buttons on the controller in CMD mode. We designed custom implementations of the raycasting and reality-based controls, but similar implementations can be found in other toolkits (Ray and Bowman, 2007).
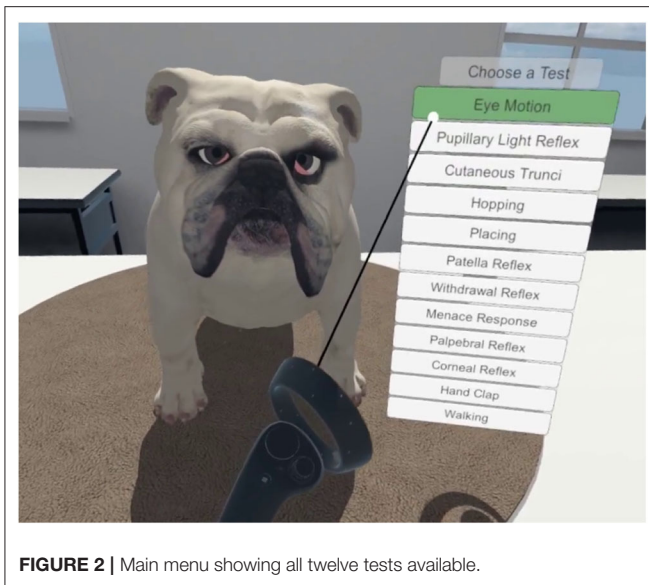
FIGURE 2 | Main menu showing all twelve tests available.

Following selection of a particular test, in CMD mode, the participant uses further menus to perform each test. After triggering the test, the participant can observe the reaction and repeat the test as necessary. In the RBI mode, the participant uses virtual hands or tools tracked by the controllers to perform each test. The exact mapping of each is reported in the following subsections. Each interface was designed through iterative design and evaluation with veterinary educators and students in formative pilot studies. We attempted to optimize each interface for consistency with the metaphor, for usability, and for the limits of the underlying simulation.

4.1. Hand Motion
Some of the tests use the location of the controller over time to generate the desired behaviors in the dog in RBI mode. These tests include the eye motion (see **Figure 3**), pupillary light reflex, menace response, hand clap, and walking tests. In all of these tests, the dog will either follow or respond to the position of the user's hand around the dog. In the hand clap test, an input is required to initiate the clap, but position of the hand still affects the response. For the CMD mode, these tests provide a list of positions or motions that are then carried out by a virtual hand. The dog only responds to the non-user virtual hands.

4.2. Touch to Interact
Other tests require the user to touch the dog with a tool or virtual hand that is tracked by the controller in RBI mode. These tests include the cutaneous trunci reflex, palpebral reflex, corneal reflex, patellar reflex, and withdrawal reflex. The trunci and withdrawal tests require a user input on the controller to initiate, but the rest are activated simply by contact of the virtual tool or hand to the appropriate location on the dog. In CMD mode, a virtual tool or hand carries out the motions required to elicit a response from the dog.

4.3. Grabbing and Drag
The last and most tangibly interactive category of RBI mode interactions includes tests that involve "grabbing" and moving a part of the dog. The withdrawal reflex and placing response (see **Figure 4**) tests make use of this interaction. In these tests, the paws of the dog are grabbed and can be moved within a bounding box to a desired location to carry out the test. The dog responds to both the movement of the paw and the final resting location. In CMD mode, each menu option initiates an animation of a virtual hand grabbing and moving a paw in the same way.

4.4. Secondary Tests and Design Compromises
While most of the identifiable symptoms present in the dog were accounted for by the twelve dedicated tests that were clearly labeled, there were also some omnipresent symptoms that did not have a specific test associated with them. These symptoms included nystagmus (uncontrolled, repetitive, eye movements), asymmetric facial droop, a visible nictitating membrane (3rd eyelid), abnormal blinking in one or both eyes, and head tilt. Proper identification of many of these symptoms represents greater awareness and attention to the whole dog rather than only observing the responses to deliberately performed tests. While the other symptoms only were visible when their test was selected, these symptoms were visible in the main menu and in every other test. For consistency with the CMD mode, we chose to only allow activation of specific reflexes while in those tests. For example, the menace response could only be elicited when in the menace response submenu and could not be activated by a similar gesture in a different test, even though a real dog would react in such a situation.

Two of the tests, withdrawal and patellar reflexes, also allowed the dog to be flipped over to test the opposite side. While performing the flip, the dog could be left in a supine (on its back) position, where any nystagmus could be observed. To make this natural in the CMD mode, the dog was rolled incrementally by pressing and holding each button. This system allows the dog to be kept in any position between the rotation endpoints, allowing the participant to observe the effect of being upside down on the direction of a potential nystagmus. When the button is released, the dog gravitates to one of the three possible orientations: either side or upside down. While a third button to roll the dog supine could have been introduced, this may have led to unnecessary leading of the user through what otherwise was an observation rather than a deliberate test.

From an interface design and simulation perspective, the hopping response was the most challenging test to map to the RBI mode, as performing the test by grabbing the opposite paw and expecting the other side of the dog to be lifted is not necessarily intuitive. However, it was deemed necessary because a two handed manipulation would have introduced more complex simulation requirements and still would have had limited fidelity to the actual test, as lifting the dog's weight is a highly physical maneuver.

5. STUDY DESIGN

We conducted a study to evaluate the usability of each interaction mode (CMD and RBI). This study was performed in preparation for the system to be deployed in the veterinary training curriculum, and it was used to determine what usability issues there were with either mode, and particularly if those usability issues impacted the utility of each mode for training and assessment. The purpose was not to determine which of the systems was better, but to improve our understanding

of each in relation to the other to inform later curricular deployment decisions.

To facilitate this comparison, we used a two-period, two-treatment crossover design. This allowed each participant to experience each interaction mode, and thus be able to comment on them both, while also allowing for an unbiased contrast between the two treatments after the first period. The two periods used virtual dogs exhibiting different abnormalities, so that the second examination was not repetitive for the student. The first dog to be examined displayed a greater facial droop on the left side, inability to blink with the left eye (also affecting palpebral, menace, and corneal reflexes), and nystagmus in both eyes. The second dog had a reduced response in the placing, hopping, patellar, and withdrawal reflexes, and displayed symptoms of ataxia (lack of control over the legs) when walking. In addition to different abnormalities, the second dog had different fur coloring to reduce the cognitive load of remembering the symptoms from the tests and avoiding confusing the two dogs. After independently performing the series of tests on the dog, the participant would note their conclusions on the patient evaluation form outside of VR.

The following hypotheses were evaluated during the study, evaluating accuracy, preference, and efficiency of each mode.

- **H1: Evaluations of patient disorders in the RBI mode will be less accurate**—The opportunity to perform tests incorrectly, or miss tests entirely, may lead to incorrect diagnoses in the RBI mode.
- **H2a: Participants will enjoy/prefer the RBI mode**—This is commonly found when comparing simulators, that more realistic interactivity is preferred by users (although this is not necessarily more useful for educational objectives) (Scerbo and Dawson, 2007).
- **H2b: Participants will find the RBI mode more difficult to use**—The CMD mode is highly consistent and should be easier for participants to learn the basic activation steps for each test.
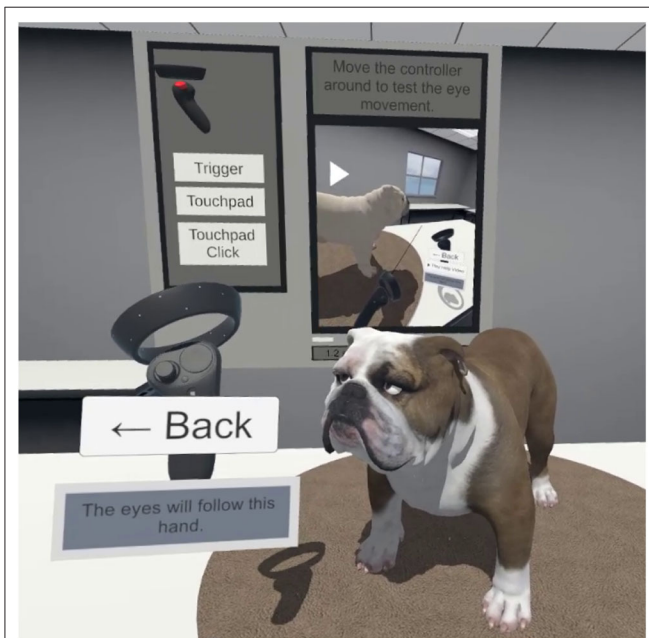- **H2c: Participants will find the RBI mode better for training**—The RBI mode offers opportunities to both practice

FIGURE 3 | The eye motion test in RBI mode. The eyes are following the controller. Also seen here is the second dog coloration.
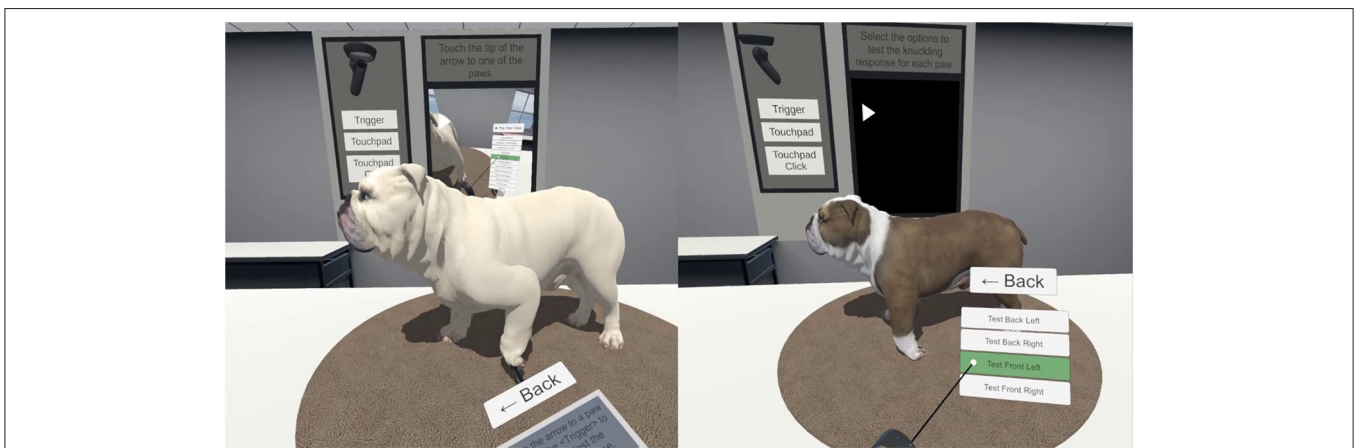

FIGURE 4 | (Left) The RBI mode for the placing test. The participant is manipulating the position of the paw directly with the controller. **(Right)** The CMD mode for the same test. This dog also has the alternate fur coloring. Note the screen on the back wall that can display a tutorial video and the model of the controller with reminders of the button names.

and learn procedures. We expected participants to recognize this and rate the RBI mode as being better for training purposes, so long as the RBI mode was easy enough to use.

- **H3: Tests will take longer to complete in the RBI mode**— Although technically, the tests could be performed faster in the RBI mode, we expected that it would take longer for users to figure out how to perform each test, and any mistakes made performing tests would add time.

5.1. Measures

5.1.1. Background Survey

The survey assessed gender (Male/Female) and age, as well as variables we anticipated may influence results: veterinary school experience (number of years), 3D game experience (5-point interval scale), confidence in performing neurological examinations (5-point interval scale), and number of neurological examinations performed (0, 1–10, 11–50, 51–100, 100+, recorded as 0, 1, 2, 3, 4).

5.1.2. Usage and Event Logs

During the course of each virtual examination, we recorded events that took place, including the start and end time of each examination test, as well as internal events within that test, such as each time the participant activated a menace test. In addition, we recorded a screen capture video using Open Broadcaster Software Studio of each participant's interactions.

5.1.3. Patient Evaluation Form

This form, completed for each dog, was modified from the standard one used at the veterinary school for neurology examinations (available upon request). It is arranged according to physical examination test findings, and we eliminated tests that we did not support in the virtual environment. It was scored according to correctness of each finding for each dog.

5.1.4. Post-experience Survey

This survey assessed opinions about each interaction system as "System 1" and "System 2" on a five-point scale (1 would indicate preference for "System 1" and 5 would indicate preference for "System 2"). The three items were general preference (which one they "liked" more), training utility (which one they thought would be better for training), and usability (which one they thought was "easier" to use).

5.2. Population and Environment

In total, 55 volunteers from our local school of veterinary medicine were recruited for the study. No incentive was provided for participation or performance. The study was approved by our Institutional Review Board, and all participants signed an informed consent document that included general objectives for the study and data collection practices.

To aid in recruitment and minimize time requirements for the students, the study was conducted at the veterinary school building, which is located several miles from our VR laboratory. A secondary effect of this was that we were able to use actual physical examination rooms for the study. Unfortunately, the same room was not available each day, and as a result the study was performed in several different rooms. The rooms were not identical, but the sizes and configurations were very similar, and

we did not expect differences in room size to have a significant effect on the data collected. In all cases, the virtual room had a larger size than the real room, so intervention of some kind was needed to prevent collisions. The Windows Mixed Reality system we were using did have a utility to display a grid in the headset denoting the real-world boundaries, but we decided to avoid the issue by monitoring participants and warning them if they neared boundaries (which occurred fewer than five times). This approach also reduced setup costs, as the boundary did not need to be measured. Prior to the start of the experiment, space calibration was performed to ensure the real space and the virtual space were aligned in such a way as to maximize the amount of real movement the participant could take that was free of obstructions.

In addition to the virtual dog, the environment contained several usability aids for participants. On the wall behind the table, there was a large virtual screen that displayed text hints and could display a short tutorial video for each of the 12 tests for both interaction systems as well as general control advice. A large model of one of the controllers was also hanging on the wall, showing which buttons were used to perform each function. Finally, the instrument used to perform each test contained a short description of its use. In practice, only the instrument help text was deemed useful by participants.

5.3. Procedure

Upon arrival, the participant provided informed consent after reading and listening to a summary of the planned study activities informing them that they would enter VR and perform a neurological examination on two different dogs with two different methods of interacting with the dog. They were not told how the interaction systems would differ or what symptoms the dogs would have. The participant then filled out the background survey. The participant was then randomly assigned to one of two orders (groups), either command-based followed by reality-based (CMD-RBI) or reality-based followed by command-based (RBI-CMD). Once the participant finished the background survey, they were given a brief introduction to the hardware, allowing them to see which buttons they would need to press on the controllers and how to adjust the headset. In preliminary tests, it was found that this introductory step helped reduce the number of times other buttons were pressed, possibly causing the simulation to be interrupted, as system menus could appear. The participant was also shown in the real world approximately where they would have to stand and where the dog would appear.

Once in VR, the participant was instructed to perform the neurological examination and identify any abnormalities. While the participant was using the system, they were given minimal outside help, but could ask questions about the interface, if necessary. After finishing the tests, the participant was asked if they wanted to go back to look at some of the tests, and after completion they filled out the patient evaluation form. The same steps, including the controller tutorial, were repeated for the second dog with the second interaction system. After completing the second patient evaluation form, the participant completed the post-experience survey comparing the two interaction modes. The ordering of the two dogs was the same for all participants in the study regardless of which interaction mode they used first.

6. RESULTS AND ANALYSIS

6.1. Background Data

All 55 participants successfully completed both examinations. None of the participants reported any motion-sickness as a result of using the system, and only one major issue occurred during the study, that the event interactions for the corneal reflex test were not recorded in the RBI mode due to an error in the software. As a result, these data were not used in comparisons. Note, the test worked correctly, but the event data were not recorded. Another minor issue was participants occasionally activating the Windows Mixed Reality menu, despite being told beforehand how to avoid this. These events were recorded but were not frequent enough to include in the analysis. All analyses were conducted using IBM SPSS.

Of the 55 participants, 50 reported as female, and five reported as male. The high ratio of females to males is common in modern veterinary school enrollment as seen in McBride et al. (2017). Of the male students, three were randomly assigned to the CMD-RBI order, and two were randomly assigned to the RBI-CMD order. As a result of the low number of male participants, we did not include gender in any analyses.

Descriptive statistics were computed for all background variables and compared between groups. No significant differences, evaluated by t-tests, were observed between groups on age (M = 24.95 years, SD = 2.69 years), VR or Gaming Experience, or confidence in performing the neurological examination. A Chi-Squared test of neurological exam experience also did not show a significant difference between groups.

6.2. Usage and Event Logs

The average time spent within each test is shown in **Figure 5**, grouped by interaction mode. As expected, the data showed that participants spent more time per test, on average, in the RBI mode (9 out of 12 tests took longer in RBI mode). A repeated measures analysis of variance, with total time as the repeated measure and order as a between subjects factor revealed a significant ($p < 0.05$) main effect of trial (dog) on average time spent per test, with the second dog taking longer (M = 380 s, SD = 87 s) than the first dog (M = 349 s, SD = 91 s). An interaction effect was also found between order and trial, with the increase in time for the CMD-RBI being significantly ($p < 0.001$) higher (Trial 1: M = 332 s, SD = 74 s; Trial 2: M = 415 s, SD = 86 s) than the RBI-CMD group (Trial 1: M = 366 s, SD = 105 s; Trial 2: M = 346 s, SD = 75 s). This confirmed that RBI mode increased time taken to perform tests.

The number of events (e.g., left front paw tested) recorded within each test also showed patterns, as seen in **Figure 5**. Similar to the time spent in each test, more events were recorded in the RBI mode for 8 out of 11 of the tests (recall that the corneal test event data were not recorded). There was a significant correlation (R = 0.63, $p < 0.01$) between the average time and the average number of events, suggesting that additional time was spent performing more tests, rather than the tests taking longer. This was most noticeable in the menace, palpebral, and patellar tests. These tests all involved striking motions, suggesting a usability

issue in RBI mode for such motions. A repeated measures analysis of variance, with total events as the repeated measure and order as a between subjects factor revealed a highly significant ($p < 0.001$) interaction effect between order and trial, indicative that the interaction mode was the primary factor affecting results, with RBI mode having more events for both trial 1 (M = 104, SD = 40.1) and trial 2 (M = 99.2, SD = 27.9) than CMD mode in trial 1 (M = 64.9, SD = 13.883) and trial 2 (M = 66.14, SD = 15.395). Note, these data show that when using the CMD mode, the number of events was roughly the same, regardless of the dog. Also note that the minimal number of events to complete all tests was 46.

6.3. Performance

Performance data showed a large differences between the difficulty of the two dogs, as shown in **Figure 6**. Participants made many more errors in evaluating the second dog (M = 5.93, SD = 2.61) than the first dog (M = 2.46, SD = 1.46), a 241% increase in errors. To evaluate the significance of this difference, a repeated measures analysis of variance was conducted, using errors made for each evaluation form as the repeated measure and order as the between subjects factor. As suspected, trial was significant at the $p < 0.001$ level. Unexpectedly, order was found to be significant ($p = 0.011$) as well. The CMD-RBI group performed significantly better on both dogs (2.10 errors, 5.22 errors) than the RBI-CMD group (2.82 errors, 6.61 errors). We speculate on the reasons for this in the discussion. No significant differences were observed in terms of completion of individual tests, with nearly 100% completion of all tests in both conditions.

6.4. Preference and Options

Preference results can be found in **Figure 7**. The majority of participants indicated that they preferred the RBI mode and that it would be better for training. The data were not normally distributed, so a one-way Chi-Square test was used to test for significance, revealing both to be significant at the $p < 0.001$ level. Similarly, the CMD mode was found to be significantly ($p < 0.026$), though not substantially, easier to use. These results were as expected.

7. DISCUSSION

Results from the study were informative, but were not entirely supportive of our hypotheses, particularly with respect to performance. H1 evaluated neurological examination performance, focusing on accuracy of responses. We hypothesized that examinations in the RBI mode would be less accurate (more errors). Our analysis, above, showed that order of interaction mode, rather than interaction mode alone was a significant factor. As a result, we could not confirm H1. However, the order effect is a potential line of future work.

One theory we have is that the order is important so that the system becomes more interactive, rather than less interactive over time. Students who experienced the first interaction with RBI mode, and then used CMD mode for the second may have been disappointed (particularly given the high preference ratings for RBI mode). By contrast, those that experienced the CMD mode
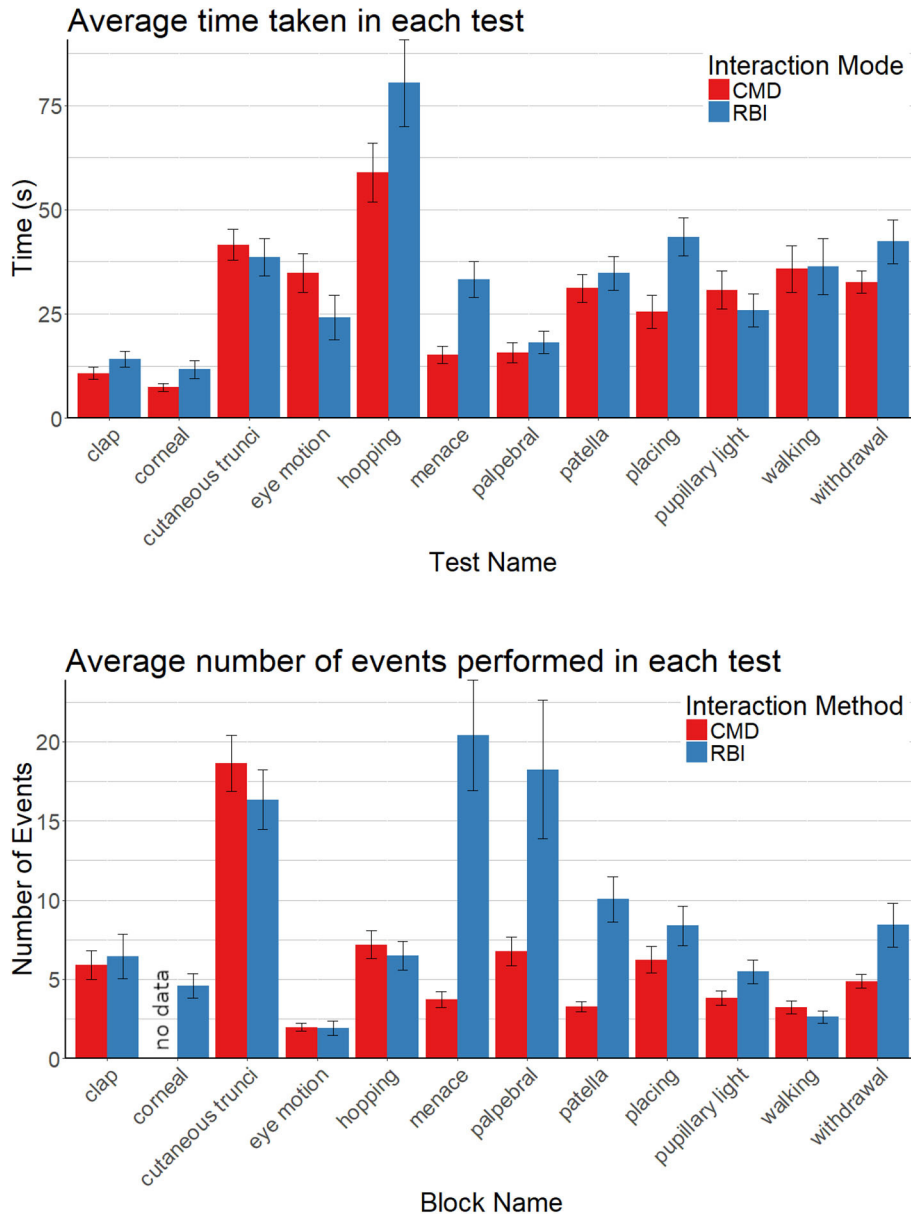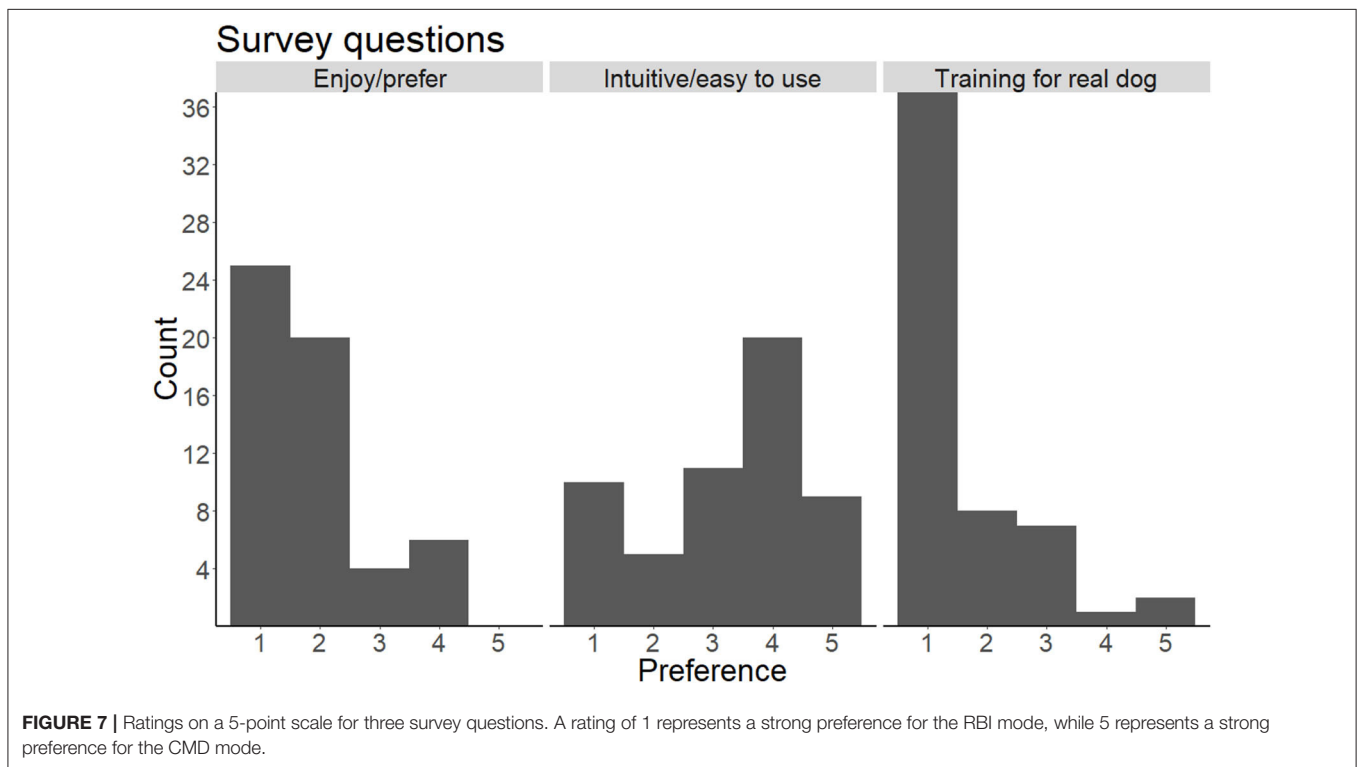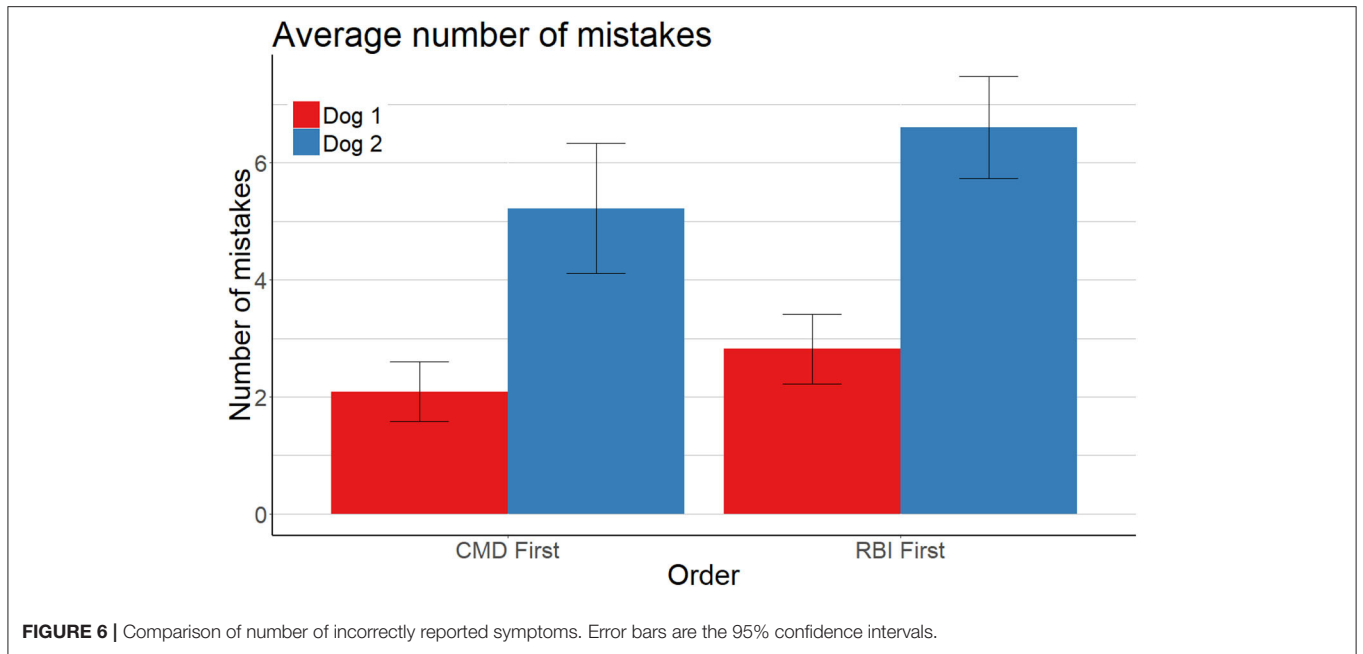
FIGURE 5 | Comparison of time spent in each test (top) and number of events completed (bottom) for the two interaction systems. Error bars are the 95% confidence intervals.

first, and then tried RBI mode were likely impressed by both. This was further supported by anecdotal comments in the CMD-RBI group in the first interaction (CMD) that they were disappointed they could not interact with their hands, even before being told that the second system would give them this ability.

A second theory is that the CMD mode is better to use first, as it shows a student how to perform an exam. By following that with RBI mode, the student would be more likely to understand how to do the tests correctly. By starting with RBI mode, the student would not have this support. We also believe that the two theories are complementary, and it is likely that a compound effect between the two caused the results.

H2a, H2b, and H2c each evaluated various participant opinions about the interaction modes and use cases. We hypothesized that participants would both enjoy the RBI mode more and find it better for training, but that they would find the CMD mode easier to use. These were all supported by analyses, with high overall preference for the RBI mode in terms of enjoyment (H2a) and usefulness for training (H2c). Results were less pronounced for ease of use (H2b), with only a statistically significant, but small, preference for CMD mode. We believe this means that, despite some observed usability issues in the RBI mode, students still did not find it very difficult to use. Overall, we can confirm all three sub hypotheses, H2a, H2b, and H2c.

FIGURE 6 | Comparison of number of incorrectly reported symptoms. Error bars are the 95% confidence intervals.

FIGURE 7 | Ratings on a 5-point scale for three survey questions. A rating of 1 represents a strong preference for the RBI mode, while 5 represents a strong preference for the CMD mode.

The final hypothesis, H3, evaluated efficiency. This hypothesis was confirmed as tests took significantly longer to complete in the RBI mode. Our reasoning, however, was likely incorrect, as the increase in time could be attributed to an increase in events, which likely was a result of usability issues rather than efficiency of performing the actions, as discussed below.

It is important to note that a significant increase in errors occurred from the first dog to the second dog. The discrepancy in number of errors between different dogs does not affect the comparison between different participants or conditions for the same dog. The greater number of errors is also not necessarily indicative of a "harder" dog to analyze, as many errors in the second dog had more parts. For example, if the participant

decided the dog had no abnormality with the withdrawal reflex, one error would be recorded for each limb. Another concern with the results could be possible negative training, since the participants seemed to perform worse on the second dog, but along with the aforementioned reasons, there is no overlap in abnormal symptoms between the two dogs, making a direct comparison between the number of errors for the first and second dogs illogical.

7.1. Observations and Usability Issues

The data indicate usability issues in the RBI mode for several tests relative to the CMD mode. These are most prominent in the tests that involved moving the hand or a tool to "strike" the dog in a location. Whereas in the CMD mode, the user could assume that the test was performed correctly, and if the dog had a response, they would observe it. In RBI mode, there was no indication if the test was performed incorrectly. The participant could have attributed this to either a neurological disorder or to not performing the test properly. As such, their only recourse was to try again to be certain that they performed the test correctly. In these cases, the participant often referred to the tutorial videos on the wall that demonstrated the correct procedure. While this lack of feedback from the dog is not inherently a problem (the real world would also elicit this behavior), the lack of haptic feedback could have made it difficult to determine if the test was performed correctly. Providing this feedback through a controller vibration (note, Windows Mixed Reality, at the time, did not support vibration through SteamVR) would be viable, or through an alternative modality (e.g., sound or visuals) could have improved the usability. In a training mode, we could also provide feedback when the reflex was activated, so that the user would know the test was performed correctly.

We do note that two tests took significantly longer to perform in the CMD mode, the eye motion test ($p < 0.001$) and the pupillary light reflex test ($p < 0.05$). These were, prior to the running the study, thought to be some of the most usable tests for RBI mode, as they do not involve touching the dog. In fact, many participants, even in CMD mode, when they had not yet used RBI mode, tried to get the dog to follow their hands. This further confirmed that the "default" expectation in VR is for RBI mode, and likely confirms that the lack of haptic feedback remains a major issue for object interaction in VR.

8. CONCLUSIONS AND FUTURE WORK

Results from the user study presented above suggest that, when using an immersive VR system, RBI is highly viable for neurological examination training in veterinary medicine education. Despite some notable usability issues likely caused by the lack of haptic feedback during physical examinations, the overwhelming preference was for reality-based interaction. However, our exploratory analysis also revealed that there may be benefit in providing the CMD mode prior to the RBI mode, as significantly fewer diagnostic errors occurred in that sequence.

This study did not evaluate any training benefits for the students, although students suggested that it has utility for

training in classrooms. We have purchased two complete VR systems to permanently house at the School of Veterinary Medicine, and based on the results of this study, intend to purchase several more to increase the possible throughput of the system. Moreover, we have adapted the CMD mode to work with desktop VR, deployable as a web application. This mode could be useful in a classroom environment, which could be followed-up by training in the RBI mode.

Lastly, our current RBI mode was artificially matched to the CMD mode by organizing the tests under menus. A more ideal application design would avoid the use of any menus, instead having the tools, such as a hemostat or cotton swab, available on a table to be picked up to use on the dog. When a tool is picked up, the test would be "activated" in the background automatically. However, this approach may have issues in the dog simulation supporting reactions to all tests simultaneously or usability issues when trying to perform different actions with the same tool (clapping and pinching with only a hand). As a result, future work will examine the possibility of higher fidelity in the neurophysiology simulation of the virtual dog to make the dog respond plausibly at all times.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Georgia Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence Teleoper. Virtual Environ.* 14, 379–393. doi: 10.1162/105474605774785235

Bogert, K., Platt, S., Haley, A., Kent, M., Edwards, G., Dookwah, H., et al. (2016). Development and use of an interactive computerized dog model to evaluate cranial nerve knowledge in veterinary students. *J. Vet. Med. Educ.* 43, 26–32. doi: 10.3138/jvme.0215-027R1

Brooks, F. P. (1999). What's real about virtual reality? *IEEE Comput. Graph. Appl.* 19, 16–27. doi: 10.1109/38.799723

Dede, C., Salzman, M. C., and Loftin, R. B. (1996). "Sciencespace: virtual realities for learning complex and abstract scientific concepts," in *Virtual Reality Annual International Symposium, 1996, Proceedings of the IEEE 1996* (Santa Clara, CA: IEEE), 246–252. doi: 10.1109/VRAIS.1996.490534

Ermak, D. M., Bower, D. W., Wood, J., Sinz, E. H., and Kothari, M. J. (2013). Incorporating simulation technology into a neurology clerkship. *J. Am. Osteop. Assoc.* 113, 628–635. doi: 10.7556/jaoa.2013.024

Issenberg, S. B., and Scalese, R. J. (2008). Simulation in health care education. *Perspect. Biol. Med.* 51, 31–46. doi: 10.1353/pbm.2008.0004

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., et al. (2007). "Reality-based interaction: unifying the new generation of interaction styles," in *CHI'07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA: ACM), 2465–2470. doi: 10.1145/1240866.1241025

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., et al. (2008). "Reality-based interaction: a framework for post-wimp interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence: ACM), 201–210. doi: 10.1145/1357054.1357089

Kaufmann, H., Schmalstieg, D., and Wagner, M. (2000). Construct3d: a virtual reality application for mathematics and geometry education. *Educ. Inform. Technol.* 5, 263–276. doi: 10.1023/A:1012049406877

McBride, S., Walker, D., Roberts, J., and Mehlhorn, J. (2017). Qualities and attributes of successful veterinary school applicants. *J. Educ. Train.* 4, 97–106. doi: 10.5296/jet.v4i2.11753

Motola, I., Devine, L. A., Chung, H. S., Sullivan, J. E., and Issenberg, S. B. (2013). Simulation in healthcare education: a best evidence practical guide. AMEE guide no. 82. *Med. Teach.* 35, e1511–e1530. doi: 10.3109/0142159X.2013.818632

Norman, G., Dore, K., and Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Med. Educ.* 46, 636–647. doi: 10.1111/j.1365-2923.2012.04243.x

Ray, A., and Bowman, D. A. (2007). "Towards a system for reusable 3d interaction techniques," in *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*

(Newport Beach, CA: ACM), 187–190. doi: 10.1145/1315184.13 15219

Reeves, B., and Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge: Cambridge University Press.

Scerbo, M. W., and Dawson, S. (2007). High fidelity, high performance? *Simul. Healthc.* 2, 224–230. doi: 10.1097/SIH.0b013e31815c25f1

Schoenherr, J. R., and Hamstra, S. J. (2017). Beyond fidelity: deconstructing the seductive simplicity of fidelity in simulator-based education in the health care professions. *Simul. Healthc.* 12, 117–123. doi: 10.1097/SIH.000000000000 0226

Stuerzlinger, W., Chapuis, O., Phillips, D., and Roussel, N. (2006). "User interface façades: towards fully adaptable user interfaces," in *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux: ACM), 309–318. doi: 10.1145/1166253.11 66301

Sutherland, I. E. (1965). "The ultimate display," in *Multimedia: From Wagner to Virtual Reality*, ed W. A. Kalenich (New York, NY: Macmillan), 506–508.

Vembar, D., Sadasivan, S., Duchowski, A., Stringfellow, P., and Gramopadhye, A. (2005). "Design of a virtual borescope: a presence study," in *Proceedings of HCI International* (Las Vegas, NV), 22–27.

Vinayagamoorthy, V., Steed, A., and Slater, M. (2005). "Building characters: lessons drawn from virtual environments," in *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop* (Stresa), 119–126.

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., and Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Appl. Ergon.* 33, 559–570. doi: 10.1016/S0003-6870(02)00039-X

Whitelock, D., Brna, P., and Holland, S. (1996). *What Is the Value of Virtual Reality for Conceptual Learning? Towards a Theoretical Framework.* Cite Report. Proceedings of the European Conference on Artificial Intelligence in Education.

Zhang, H., Starke, S., Komura, T., and Saito, J. (2018). "Mode-adaptive neural networks for quadruped motion control," in *ACM Transactions on Graphics/SIGGRAPH 2018.* Vancouver. doi: 10.1145/3197517.3201366