To appear in 2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

The Impact of Performance-Specific Feedback from a Virtual Coach in a Virtual Reality Exercise Application

Andrea Zimmerer*
HCI Group, University of Würzburg

Lydia Bartels†
HCI Group, University of Würzburg

Marc Erich Latoschik‡
HCI Group, University of Würzburg

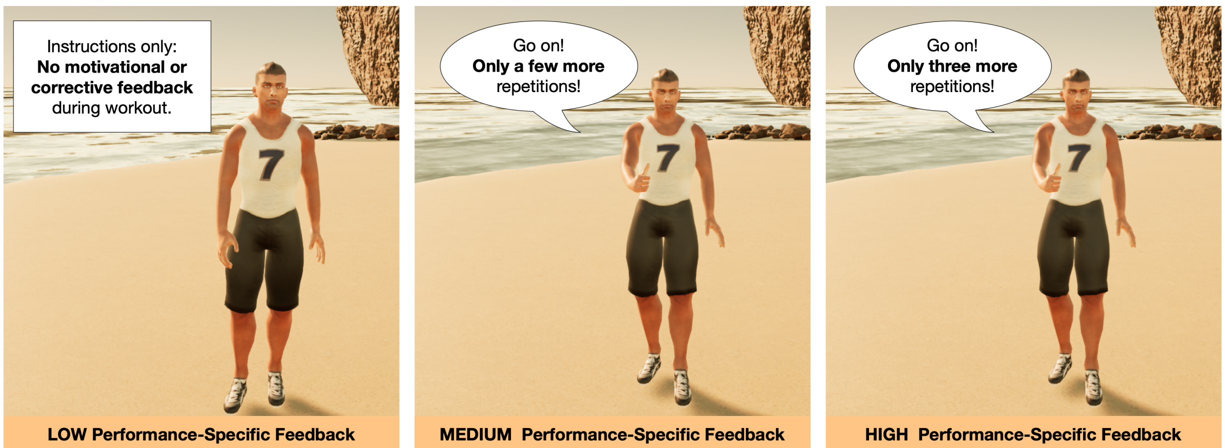

Figure 1: The virtual coach guided users through their workout in the VR exercise application by talking to them. Depending on the experimental condition he provided feedback with low, medium or high performance-specificity. Text annotations in this figure are for illustration purposes only and were not part of the VR exposure.

Abstract

Virtual reality (VR) exercise applications are promising tools, e.g., for at-home training and rehabilitation. However, existing applications vary significantly in key design choices such as environments, embodiment, and virtual coaching, making it difficult to derive clear design guidelines. A prominent design choice is the use of embodied virtual coaches, which guide user interaction and provide feedback. In a user study with 76 participants, we investigated how different levels of performance specificity in feedback from an embodied virtual coach affect intermediate factors, such as VR experience, motivation, and coach perception. Participants performed lower-body movement exercises, i.e., Leg Raises and Knee Extensions, commonly used in knee rehabilitation. We found that highly performance-specific feedback led to higher scores compared to medium specificity for perceived realism, as well as the anthropomorphism and sympathy of the virtual coach, but did not affect motivation. Based on our findings, we propose the design suggestion to include precise, performance-specific details when creating feedback for a virtual coach. We observed a descriptive pattern of higher scores in the low specificity condition compared to the medium condition on most measures, which raises the possibility that less specific feedback may, in some cases, be perceived more positively than moderately specific feedback. These findings provide valuable insights into how design choices impact relevant intermediate factors that are crucial for maximizing both workout effectiveness and the quality of the virtual coaching experience.

Index Terms: Agents, Virtual Reality, Exercise Systems.

*e-mail: andrea.zimmerer@uni-wuerzburg.de
†e-mail: lydia.bartels96@web.de
‡e-mail: marc.latoschik@uni-wuerzburg.de

1 Introduction

VR exercise systems offer exercise support across various domains, e.g., general fitness, professional sports coaching, and rehabilitation at home [9, 28, 45]. Research indicates that VR exercise systems can effectively enhance performance [28, 45]. However, the systems used vary significantly regarding substantial design choices, e.g., in terms of the technology employed and, consequently, their level of immersiveness, the exercise environments, user embodiment, feedback strategies, and more. This variability makes it difficult to assess the impact of individual design choices and their combinations on the application's success, complicating the derivation of clear design guidelines. As a result, the specific characteristics that contribute to optimal effectiveness remain an ongoing area of investigation. A key first step is understanding how design choices influence intermediate factors that ultimately contribute to the application's desired outcome [2, 45]. For example, IJsselsteijn et al. [30] found that immersion (design choice) positively influences motivation (intermediate factor), which, in turn, increases workout adherence, leading to greater workout effectiveness.

Most digitally enhanced training systems provide some form of virtual coaching. The design space for virtual coaching is vast, with almost every design decision opening up new research avenues [61]. A common design choice is the use of embodied virtual coaches [22], which serve as analogs to real-life coaches and add the benefit of a social component. A human-like embodiment of the coach has additional advantages: the virtual coach can demonstrate exercises and even exercise alongside the user, potentially enhancing the workout process by activating the mirror neurons [27].

Embodied virtual coaches typically introduce exercises, provide corrective feedback, and motivate users to stay engaged [22]. The most basic form of a virtual coach offers general guidance, potentially including user-agnostic feedback, which is easy-to-implement and cost-effective. However, this leaves much of the technology's potential untapped [13]. Since users in most VR exercise systems already wear sensors that collect, for example, movement and physiological data, more advanced exercise systems can leverage the rise of large language models and artificial intelligence to optimally tailor the experience to the user's performance, preferences, and needs [1, 10, 29, 57]. This allows the coach to provide specific feedback on a user's performance. However, performance-specific feedback is not yet standard practice, even in systems explicitly designed with wearable technologies [56]. Consequently, little is known about the impact of performance-specific feedback provided by embodied virtual coaches in VR exercise applications.

To address this research gap, we conducted a user study in which an embodied virtual coach provided verbal feedback in a highly immersive VR exercise application. Participants performed lower-body movement exercises that are commonly used in knee rehabilitation: Leg Raises and Knee Extensions. A virtual coach gave feedback with low, medium, or high performance specificity. We assessed participants' perceptions of the VR experience, motivation, and perception of the virtual coach. Our findings show that feedback with high performance specificity led to higher perceived realism and made the virtual coach seem more anthropomorphic and sympathetic compared to the medium condition. We observed a non-significant pattern of higher scores in the low specificity condition compared to the medium condition, which points to the possibility that less specific feedback may be perceived more favorably under yet to be defined circumstances. Motivation did not differ significantly across conditions. Based on our findings, we derive the design suggestion to include performance-specific details when designing feedback for a virtual coach in a VR exercise application. Our results provide valuable insights into how basic design choices influence intermediate factors crucial for maximizing both coaching effectiveness and the quality of the virtual coaching experience.

2 Related Work

Research uses various terms to describe adaptations of VR systems to users, e.g., personalization, customization, or individualization [14, 36, 42, 59]. Regardless of the specific term, adaptations generally fall into one of two categories: Adaptations initiated by users and adaptations performed by the system based on available user data. In our work, we focus on the latter type, referring to it as a *system-driven adaptation*. Specifically, we investigate how feedback, adapted in terms of *performance specificity* and delivered by an embodied virtual coach during a sports workout, impacts intermediate factors in a VR exercise application. We concentrate on three key intermediate factors: the overall perception of the VR experience, motivation, and the perception of the virtual coach.

Perception of the VR Experience Two central concepts regarding the perception of VR experiences are *Presence* and *Immersion* [51]. Presence refers to the feeling of truly being in a virtual environment [51], which is promoted by high levels of immersion [51, 54, 59]. Slater et al. [52, 53] described immersion as a characteristic of the medium itself, with higher-immersion media engaging more senses and enabling more valid actions than low-immersion systems [51, 52]. IJsselsteijn et al. [30] found that high immersion promotes both presence and motivation in a cycling task, highlighting the relevance of high-immersive systems for VR exercise applications. Research indicates that system adaptations to the user can potentially enhance this immersive experience. For example, it has been shown that a user's opportunity to customize avatars in video games increased perceived immersiveness and player satisfaction [55]. In educational settings, customizable avatars have been linked to a greater sense of presence and engagement [14]. Similarly, Pardini et al. [42] found positive effects on the VR experience when users could adapt aspects of the environment, such as weather or landscape, in a relaxation application. However, these examples involved user-driven adaptations. Empirical research on the perception of system-driven adaptations remains sparse. Marougkas et al. [41] introduced content adaptations by both users and systems as an important element in engaging VR experiences for educational applications. However, there is a lack of empirical studies examining the impact of system-driven adaptations of feedback provided by virtual coaches in VR exercise applications on the perception of the VR experience.

Motivation A commonly predicted advantage of VR exercise systems is their potential to enhance motivation by providing engaging and enjoyable experiences [28]. Motivation is the drive to act in order to achieve a goal perceived as rewarding [50]. Key components of motivation are perceived competence, autonomy, and relatedness with others [16]. Motivation is central in sports, as it is connected to an athlete's performance [58] and it is particularly important in physical rehabilitation, where success of the rehabilitation depends on patient compliance and their motivation to exercise at home [63]. Research has shown that feedback influences motivation. Deci and Ryan [17] state that positive feedback that enhances perceived competence and autonomy can improve intrinsic motivation. Similarly, Mageau and Vallerand [38] emphasize that positive, non-controlling feedback that fosters competence, refers to aspects the athlete can actually control, and sets challenging yet achievable goals may facilitate intrinsic motivation. Given the individual differences among athletes in terms of their performance, abilities, preferences, and needs, the adaptation of feedback is a necessity for boosting motivation [1, 18]. While several empirical studies have explored the relationship between feedback and motivation in the context of embodied virtual coaches, most have focused on educational settings [34]. However, less is known about how established theories of feedback and motivation can be applied to virtual coaches in VR exercise applications to optimize user motivation.

Perception of the Virtual Coach In real-world coaching, a good relationship between coach and athlete is essential for effective coaching, fostering trust, motivation, and better outcomes [31, 38]. Research by Garcia et al. [25] suggests that how athletes perceive their coach influences their reaction to feedback. Their study found that positive feedback from a coach perceived as less competent improved motivation and performance, whereas feedback from a more competent coach had no significant effect compared to negative or no feedback. Similarly, research finds that the relationship between users and virtual coaches is essential for effective coaching [8, 12]. Various design choices influence how users perceive virtual coaches, e.g., external factors such as appearance, voice, and word choice, but also internal factors such as different feedback strategies [4, 7, 8, 34, 60]. Overall, empirical research has concentrated more on the non-verbal than on the verbal behavior of virtual agents. In a review article on the effects of intelligent agents' multimodal behavior on their perception, Etienne et al. [19] list several effects of verbal behavior. According to their findings, verbal behavior may affect the virtual agent's perceived friendliness, dominance, competence, trustworthiness, anthropomorphism, liveliness, and likability. The verbal behaviors included in the reviewed studies were friendly cues such as informal language and shorter sentences, interruption handling times, pitch, speech rate, voice realism, and choice of pronouns. One study included in the review investigated the effects of adapting verbal behavior: Biancardi et al. [11] found that their virtual agent was perceived as more friendly when it adapted its verbal behavior according to the user's impressions. In their review article, Etienne et al. [19] state that most of the reviewed studies used low-immersive systems. Additionally, empirical research on virtual coaches predominantly concentrates on an educational context. Consequently, little is known about the effects of system-driven adaptation of feedback provided by an embodied virtual coach in a high-immersive VR exercise application. However, since high-immersive systems potentially enhance emotional reactions [23] and design choices are context-sensitive [4, 44] it is crucial to investigate the effects of feedback adaptation in terms of performance specificity in this realm.

Table 1: Protocol for the verbal feedback of the virtual coach including all provided feedback sentences for the medium and high performance specificity groups with the main differences written in bold font. Steps 2 to 9 were repeated for the second exercise. The Introduction (1), Exercise Instruction (2), Instructions to Pause (6, 10), and the Goodbye (12) were identical across all three conditions.

Feedback Protocol	Low	Medium	High
(1) Introduction			
(2) Exercise Instruction			
(3) Motivational feedb. after 7 reps	none	"Great, keep it up! You have already completed **a large part** of the exercise!"	"Great, keep it up! You have already completed **half of** the exercise!"
(4) Motivational feedb. after 10 reps	none	"Go on, only **a few more** repetitions! Let's go!"	"Go on, only **three more** repetitions! Let's go!"
(5) Corrective feedback after 13 reps	none	"Pay attention to an even movement of your leg."	"Pay **even more** attention to an even movement of your leg."
(6) Instruction to Pause			
(7) Corrective feedback after 7 reps	none	"You are doing well, a good pace in the execution of the exercise is important."	"You are doing well, but try to do the exercise **a little slower/faster**."
(8) Motivational feedb. after 10 reps	none	"Go on, only **a few more** repetitions! Let's go!"	"Go on, only **three more** repetitions! Let's go!"
(9) Corrective feedback after 13 reps	none	"Pay attention to an even movement of your leg."	"**That was better**, but **continue** to pay attention to an even movement of your leg."
(10) Instruction to Pause			
(11) Overall feedback on time spent	none	"You completed your workout in **a good time**."	"You completed your workout in **a good time under X minutes**."
(12) Goodbye			

2.1 Summary and Hypotheses

Firstly, regarding the *perception of the VR experience*, research on user-driven adaptations of employed avatars and environments in VR applications suggests positive effects [14, 42, 55]. Based on these findings, we hypothesize that a system-driven adaptation of verbal feedback to the user's specific performance will also positively impact the VR experience. Secondly, research on *motivation* suggests that feedback that responds to an athlete's particular situation can increase motivation [16, 17, 58]. We therefore expect motivation to be higher when the verbal feedback from a virtual coach is more performance-specific. Third, feedback adaptation directly impacts what the virtual coach is saying, thereby potentially changing users' *perception of the virtual coach* [19, 34]. Based on these findings, we formulate three hypotheses:

- **H1** Performance-specific verbal feedback improves the perception of the VR experience.
- **H2** Performance-specific verbal feedback increases motivation.
- **H3** Performance-specific verbal feedback changes how users perceive the virtual coach that provides the feedback.

3 Study

To test our hypotheses, we conducted a user study. The study design adhered to the ethical principles of the Declaration of Helsinki. The institution's responsible ethics committee approved the study.

3.1 Design

In our user study, participants completed movement exercises in a VR application. To avoid undesirable effects due to fatigue, the experiment followed a between-subjects design with one independent variable (IV): the performance specificity of feedback with three levels – low, medium, and high specificity. On the lowest level, the coach simply guides the user through the workout, providing basic instructions without any additional feedback. Performance specificity of the feedback is low as it only takes into account whether the exercise is completed or not. On the medium level, users receive generic, performance-related feedback, which does not include specific data but offers general advice on how to improve exercise performance. At the highest level, the coach delivers feedback that includes data specific to the user's performance, detailing progress with concrete values, e.g., completed repetitions and time spent exercising, and offering tailored suggestions for improvement, e.g., to move slower or faster. Table 1 shows the verbal feedback in all conditions. Dependent variables were the perception of the VR experience, motivation, and the perception of the virtual coach.

3.2 Material

This section describes our stimulus material, i.e., the virtual coach, the VR exercise application, the employed measurements, and soft- and hardware used. Our VR exercise application builds on our earlier work [2], which we adapted and extended to meet the requirements of this study. The video included in the supplementary material shows the VR exercise application from a first-person view.

3.2.1 Virtual Coach

Character Model Figure 1 shows the virtual coach in our VR exercise application. Inspired by real-world coaching scenarios, we selected a virtual human to serve as the coach. This human-like embodiment offers additional benefits, as the virtual coach can demonstrate exercises and engage in the activity alongside the user, potentially enhancing the training process by activating the mirror neurons [27]. Based on data regarding the gender distribution of personal trainers in the country where the study was conducted, which showed that most sports coaches are male [21], we chose a male virtual agent. He has an athletic physique and wears sportswear. The 3D character model was created using Autodesk Generator 4.

Verbal Feedback In line with the review article by Gago-Masague et al. [22], our virtual coach had three main tasks: Introducing the exercises and providing corrective as well as motivational feedback. Feedback types observed in sports are numerous and often not clearly distinct, with the most common forms being simple praise or encouragement [15]. However, research on motivation and performance in sports provides guidance on the design of effective feedback. According to Deci and Ryan [17], positive feedback that promotes athlete competence and autonomy potentially improves intrinsic motivation. Similarly, Mageau and Vallerand [38] describe that positive feedback that is not controlling but promoting competence may facilitate intrinsic motivation

Figure 2: Left: 1st-person view of a user embodied with the female avatar during the Leg Raises exercise. Right: 1st-person view of a user embodied with the male avatar during the Knee Extensions exercise.

if it promotes feelings of competence, refers to aspects the athlete can actually control, and sets challenging but still achievable goals. Following these guidelines, we scripted several voice lines for the virtual coach. Table 1 gives an overview of the whole interaction and lists all the verbal motivational and corrective feedback provided by the coach.

At the beginning of the VR exposure, the coach introduced himself, stating his name and explaining that he would guide the user through the workout. He proceeded with an explanation of the first exercise. The introduction and explanation were consistent across conditions. During the exercises (see subsubsection 3.2.2 for a full description), the coach provided feedback after 7, 10, and 13 repetitions. Two types of feedback were given: motivational feedback, which aimed to encourage the user, and corrective feedback, which focused on improving exercise performance. Performance specificity of the motivational feedback targeted the number of completed repetitions (see Table 1, lines 3-4, and 8). Performance specificity of the corrective feedback was based on additional time measurements, i.e., time per repetition and overall time for the exercise. The coach either commented on the movement speed or the consistency of the movement. For feedback related to movement speed, the system compared the elapsed time to a fixed reference value (Leg Raises: 15s, Knee Extensions: 20s). If the elapsed time exceeded/fell below this threshold, the coach provided feedback encouraging a slightly faster/slower pace (see Table 1, line 7). Movement consistency was assessed by calculating the variance in repetition durations across the 13 repetitions. This variance was then compared to a predefined threshold of 0.5s. If the variance was higher than the threshold, indicating irregular execution speed, the coach provided feedback to promote a more uniform movement pattern (see Table 1, lines 5 and 9). Between exercise rounds as well as between the first and the second exercise, the coach gave instructions on when to pause and when to resume. At the end of the workout, the coach gave praise in all conditions, followed by feedback regarding the total time spent in the medium and high specificity condition (see Table 1, line 11).

The coach's behavior was implemented using a feedback controller blueprint developed in Unreal Engine. The blueprint triggers distinct events based on the user's progress in the workout sequence, initiating the required calculations. We recorded the voice lines for the coach's explanations and feedback using the sound software Audacity5. The voice lines were spoken by a male in the same age range as the virtual character model. We tested and improved all thresholds and the initial sequences in several pilot runs.

Animation Following the review by Etienne et al. [19] that underlines the importance of combining verbal and non-verbal behavior for convincing virtual agents, we created suitable animations to go with the voice lines of our virtual coach. Specifically, idle movements, conversational gestures, the exercise movements for the coach to display during the instructions, as well as motivational gestures, e.g., a thumbs-up and an encouraging motion where the hands are swung through the air in a forward, sweeping motion. Whenever the coach spoke, the speech was accompanied by appropriate movements. For identical voice lines across conditions, the animations were also the same. When the coach was silent, he performed idle movements. To make the animations appear as human as possible, we recorded movements of a real person using the motion capture software Ikinema Orion. The hardware used for this purpose included the HTC Vive Pro head-mounted display, two hand controllers, and three additional Vive trackers (v 1.0), one on each foot and one on the back. To correct inaccuracies of the tracking system used, we post-processed the animations manually in the Unreal Engine. Specifically, the position and rotation of the individual bones of the model were revised and adjusted frame by frame. In this step, we also added finger movements and facial animations, i.e., random blinking and mouth movements that visually matched the text spoken by the virtual coach. The recorded animation sequences ranged between two and 40 seconds in length.

3.2.2 Virtual Environment, Embodiment, and Exercises

Figure 2 shows the first-person view of a user exercising in our VR application. The user's avatar sat on a virtual mat. In front of them was a virtual mirror. Mirrors are a common tool in VR applications to induce a Sense of Embodiment [32, 47, 59] but also in sports for observing one's own motion. Next to the mirror was a sign showing the required and completed number of repetitions and an angle that represents the position of the user's leg. To the right of the user's avatar stood the virtual coach who was also visible in the mirror. We made sure that the user could comfortably observe the virtual coach from their position. The virtual environment depicted a beach by the sea, with nature sounds included, and was designed to convey a positive atmosphere [2]. The user was represented by a full-body avatar to support observational learning during the exercises [2, 62]. The avatar was either female or male, matching the gender expression of the participant. We used the character models included with Ikinema Orion but with adjusted textures to look more like typical sports clothes. We selected two simple lower body movement exercises inspired by exercises commonly prescribed in

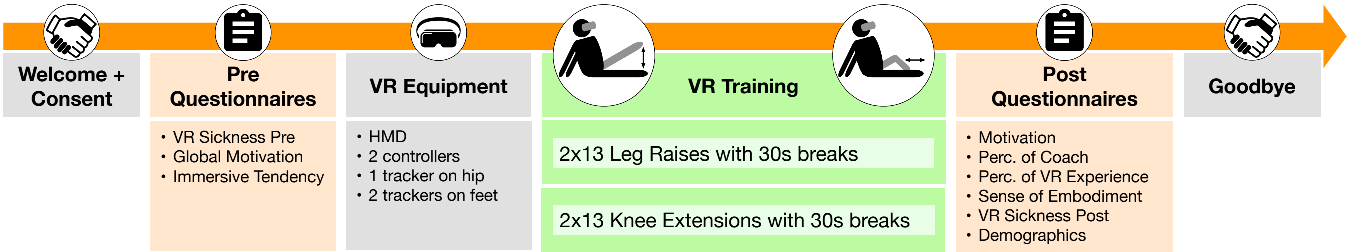

Figure 3: Procedure of the user study.

knee rehabilitation [2]. To simulate a realistic VR exercise application, both exercises incorporate basic gamification elements. The first exercise was Leg Raises. Participants had to raise their foot to a predefined position above the ground while keeping their leg extended and then lower the leg again. Gamification for this exercise was a virtual ball that users moved up and down on the tip of their foot as they performed the leg movement. The left image in Figure 2 shows the 1st person view within this exercise. The second exercise was Knee Extensions (see Figure 2, right). Participants had to bend and extend their leg, pushing and pulling a virtual box attached to their moving foot. Both the box and the ball provided visual feedback by briefly flashing green whenever a repetition was successfully completed. To track repetitions, the system calculated the angle between the leg and the ground for the Leg Raises and the knee joint angle for the Knee Extensions. Once the exercise-specific predefined threshold was reached, the movement was classified as a completed repetition.

3.2.3 Soft- and Hardware

The VR exercise application was implemented using Unreal Engine 4.17. For tracking the user's movements, we used the HTC Vive Pro with two hand controllers and three additional VIVE trackers (v. 1.0). Two of them were placed on the participants' feet and one on the lower back. IKinema Orion (v. 0.93) was used to represent the user's movements in the virtual environment. It uses six tracking points and inverse kinematics to simulate full-body tracking. The application ran on a VR-capable PC consisting of an Intel Xeon E3-1230 v5, an Nvidia GeForce GTX 980 Ti, and 16 GB RAM.

3.2.4 Measurements

Dependent Variables We used the Presence Questionnaire (PQ) of Witmer and Singer [64, 65] in the revised version of Robillard et al. [46] to assess the subjective VR experience. This version differentiates the subscales *realism*, *possibility to act*, *quality of interface*, *self-evaluation of the performance*, *sound*, *haptic*, and *possibility to examine*. We excluded the last two from our evaluation since the system does not use haptic feedback and the tasks did not allow users to examine the virtual environment very much. The item scale ranged from 1 to 7 where 7 indicates stronger feelings related to the VR experience.

To measure the participant's motivation we used five subscales of the Intrinsic Motivation Inventory (IMI) [20]: enjoyment, perceived competence, effort/importance, pressure/tension, and relatedness. We did not assess the usefulness subscale, as participants received credit points for participation, which could confound perceived usefulness. Perceived choice was also excluded, since participants only indirectly chose the task by opting into the study. The item scale ranged from 1 to 7 with 7 indicating highest motivation.

We assessed the participant's perception of the virtual coach with the Godspeed Questionnaire (GQ) [3]. This questionnaire was originally developed for the evaluation of social robots but was also successfully deployed for the evaluation of virtual agents in the past [37, 49]. We used the subscales Anthropomorphism, Sympathy, and Perceived Intelligence and excluded the subscales Animacy and Perceived Safety since they are more suitable for the evaluation of robots than virtual agents. Each subscale consists of five semantic differentials with five levels in between the two poles.

Control Measures Our VR application focuses on physical movement, making the body and its motion a central component. To account for potential confounding factors, we recorded the Sense of Embodiment (SoE) as a control variable to ensure that variations in the coach's feedback did not influence participants' perceptions of their virtual embodiment. The SoE was measured with the Virtual Embodiment Questionnaire (VEQ) of Roth and Latoschik [47, 48]. Specifically, we used a revised version of the Alpha IVBO [48] with the original structure but the revised wording of the latest version of the VEQ [47] which was not yet available at the time of study conduction. The utilized questionnaire consists of the three subscales Ownership, Agency, and Change. The item scale ranged from 1 to 7 where 7 indicates a strong SoE.

An additional prominent confounding factor for VR experiences is VR sickness. We measured the participants' VR sickness symptoms before and after the experiment using the VR sickness questionnaire (VRSQ) of Kim et al. [33], which consists of 9 items ranging from 0 to 3 resulting in a VR sickness score from 0 to 100.

Demographics and User Traits We asked participants for several demographics, i.e., sex, age, prior VR experience, sports habits, visual impairment corrections, and native language. In addition, we checked for differences between the groups regarding user traits that are important in context of the study, i.e., Immersive Tendencies and Global Motivation. We used the Immersive Tendencies Questionnaire (ITQ) [65] to measure a person's inherent capability to become immersed in a narrative [51]. To control for an unwanted bias due to differences between the experimental groups regarding their general motivation participants answered the Global Motivation Scale (GMS) [26, 43]. Both questionnaires' items range from 1 to 7 where 7 means high immersive tendency and motivation.

3.3 Procedure

Figure 3 outlines the user study procedure. Participants were randomly assigned to one of three experimental conditions, gave informed consent, and completed pre-questionnaires (VRSQ, GMS, ITQ). After an introduction and safety briefing regarding the VR setup, the experimenter equipped them with the motion trackers, controllers, and HMD, and completed calibration for motion tracking. Participants then took a seat on an exercise mat. In VR, the virtual coach introduced himself and guided participants through both exercises (Leg Raises and Knee Extensions), each performed in two sets of 13 repetitions, with 30 seconds of rest after each set. Feedback was provided according to condition (Table 1). After final feedback and farewell from the coach, participants removed the HMD and completed post-questionnaires (IMI, Godspeed, PQ, VEQ, VRSQ, demographics). Finally, participants removed the remaining trackers and were thanked and bid farewell.

3.4 Participants

A total of $N = 83$ students participated in the study. We excluded four data sets due to technical difficulties, i.e., disturbances in the tracking. As strong VR sickness symptoms may compromise participants potentially affecting result validity, we performed an outlier analysis of pre- and post-experiment VR sickness scores using boxplots. Following this analysis, three participants were excluded because they were outliers regarding their VR sickness score before the experiment ($n = 1$ $score = 55.83$), after ($n = 1$, $score = 35$) or both ($n = 1$, $pre - score = 39.17$ and $post - score = 37.5$). The remaining sample of $N = 76$ participants was $M(SD) = 21.26(2.55)$ years of age. 69.7% were female, 30.3% male. They were randomly assigned to the three experimental groups. Table 2 lists the key demographics of the sample by group. We used a Chi-square tests to examine whether there were any significant differences in the distribution of the key characteristics, i.e., sex, prior VR experience, and physical activity level, across groups. We found no significant association between the condition and the participants' reported sex, $\chi^2(2) = 1.07$, $p = .605$, their prior VR experience in hours, $\chi^2(10) = 10.45$, $p = .549$, their prior VR experience in number of exposures, $\chi^2(10) = 10.66$, $p = .564$, and the number of weekly hours they are doing sports, $\chi^2(10) = 7.55$, $p = .799$. In all groups, between 20 and 30 % had less than 1 hour of VR experience and more than 70 % had 5 hours of prior VR experience at a maximum. Also, in all groups, between 20 and 30 % stated to do less than 1 hour of sports activities every week and more than 70 % reported to do 5 hours of sports activities per week at a maximum. The participants did not differ significantly between conditions regarding their immersive tendencies, $F(2, 29.75) = 0.68, p = .515$ or their general motivation, $F(2, 29.66) = 0.29, p = .749$.

Table 2: Key demographics by experimental group. Values for sex denote the number of female/male participants in the group.

Characteristic	Low	Medium	High
Number of participants	25	24	27
Mean age (*SD*)	21.40(2.60)	20.63(1.61)	21.70(3.01)
Sex (f/m)	19/6	15/9	19/8

4 Results

The analysis was conducted with IBM SPSS Statistics 29 and R 4.1.2. Due to several normality violations, we calculated robust ANOVAs for independent samples as described by Mair and Wilcox [39], i.e., we used the *t1way* function of the WRS2 package which uses 20% trimmed means. As suggested by the authors we report the robust effect size $\hat{\xi}$. Values of .10., .30, and .50 correspond to small, medium, and large effects [39]. Post-hoc tests are pairwise comparisons calculated with the *lincon* function of the WRS2 package which returns pairwise trimmed mean differences ($\hat{\Psi}$). See Table 3 for the descriptive statistics as well as a summary of the robust ANOVA test statistics.

4.1 Perception of the VR Experience

Figure 4 shows the results for the Presence Questionnaire. We found a significant difference for perceived realism ($p < .05$, $\hat{\xi} = .38$). Perceived realism was significantly higher in the high specificity than the medium specificity condition, $\hat{\Psi} = -.52$, $95\% CI[-1.04, -0.01]$, $p = .048$. There were no significant differences regarding the other subscales.

4.2 Motivation

We found no significant differences regarding the subscales of the IMI. However, the relatedness subscale showed a non-significant trend ($p = .053, \hat{\xi} = .47$). Participants in the condition with high specificity tended to report higher relatedness, than participants who received medium, $\hat{\Psi} = -0.74, 95\% CI[-1.57, 0.10], p = .068$, or low specificity, $\hat{\Psi} = -0.66, 95\% CI[-1.39, 0.06], p = .068$.

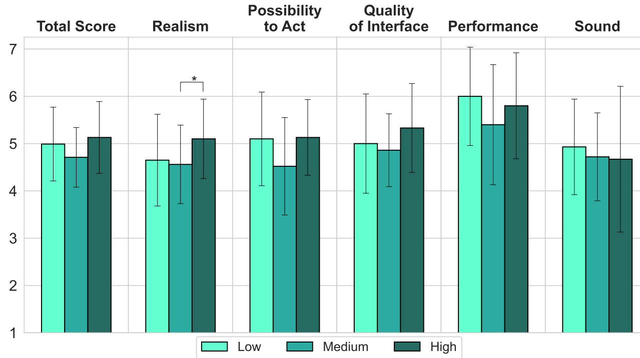

Figure 4: Mean ratings on the subscales of the Presence Questionnaire. Error bars represent standard deviations. $* < .05$

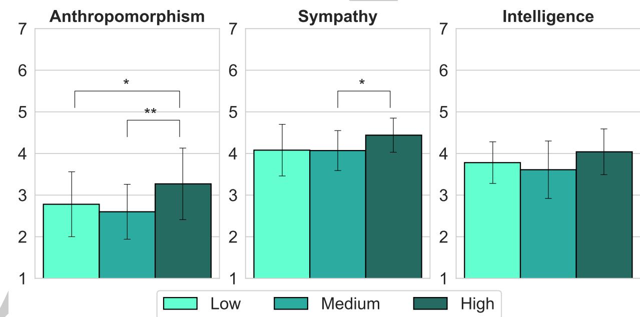

Figure 5: Mean ratings on the subscales of the Godspeed Questionnaire. Error bars represent standard deviations. $* < .05, ** < .01$

4.3 Perception of the Virtual Coach

Figure 5 shows the results for the Godspeed Questionnaire. We found a significant difference regarding the virtual coach's perceived anthropomorphism ($p < .01, \hat{\xi} = .5$) and sympathy ($p < .05, \hat{\xi} = .44$). Post-hoc tests revealed that participants who received highly performance-specific feedback found the coach to be more anthropomorph, than when they received medium, $\hat{\Psi} = -0.89, 95\% CI[-1.53, -0.26], p = .004$, or low specificity, $\hat{\Psi} = -0.75, 95\% CI[-1.42, -0.08], p = .017$. Perceived sympathy was higher with high specificity than with medium specificity, $\hat{\Psi} = -0.35, 95\% CI[-0.66, -0.04], p = .026$. There was no significant effect on the coach's perceived intelligence.

4.4 Control Measures

Our control measures showed no significant influence of the independent variable. We did not find any significant differences regarding the embodiment measurements body ownership, agency, and change. Participants' VR sickness symptoms did not differ significantly between conditions neither before, nor after the experiment. Paired samples t-tests did not show any significant difference between the pre and post measurements, neither for the low, $t(24) = -0.16, p = .874, d = -.03$, the medium, $t(23) = -1.70, p = .102, d = -.35$, nor the high specificity condition, $t(26) = 0.27, p = .793, d = .05$.

Table 3: Means and standard deviations for all dependent variables split into the experimental groups low, medium, and high performance specificity of the provided feedback, and test statistic for the calculated robust ANOVAs. $* < .05$, $** < .01$

	Low ($N = 25$)	**Medium** ($N = 24$)	**High** ($N = 27$)				
Measure	$M(SD)$	$M(SD)$	$M(SD)$	F	df	p	$\hat{\xi}$[95%CI]
Perc. of VR Experience							
Presence (total)	4.99(0.78)	4.71(0.63)	5.13(0.76)	2.91	(2,29.63)	.070	0.37[0.15,0.56]
Realism	4.65(0.97)	4.56(0.83)	5.07(0.84)	3.38	(2,29.48)	.048*	0.38[0.07,0.67]
Possibility to act	5.10(0.99)	4.52(1.03)	5.10(0.80)	1.81	(2,29.42)	.182	0.35[0.10,0.60]
Quality of interface	5.00(1.05)	4.86(0.77)	5.33(0.94)	2.76	(2,29.53)	.080	0.37[0.10,0.61]
Performance	6.00(1.04)	5.40(1.27)	5.80(1.12)	1.08	(2,29.92)	.354	0.25[0.03,0.49]
Sound	4.93(1.01)	4.72(0.93)	4.67(1.54)	0.33	(2,28.17)	.724	0.19[0.05,0.44]
Motivation							
Enjoyment	3.77(1.11)	3.56(1.24)	4.20(1.26)	0.92	(2,28.82)	.411	0.27[0.07,0.55]
Perceived Competence	3.87(1.25)	3.79(0.88)	3.75(0.90)	0.27	(2,28.63)	.763	0.19[0.03,0.46]
Effort/Importance	4.79(0.60)	5.05(0.82)	5.07(0.90)	0.80	(2,29.80)	.458	0.23[0.03,0.45]
Pressure/Tension	3.41(1.05)	3.44(1.03)	3.53(1.29)	0.14	(2,29.35)	.869	0.19[0.03,0.42]
Relatedness	3.97(1.03)	3.71(1.10)	4.64(1.14)	3.25	(2,29.59)	.053	0.47[0.13,0.69]
Perc. of Virtual Coach							
Anthropomorphism	2.78(0.78)	2.60(0.66)	3.27(0.86)	6.50	(2,29.66)	.005**	0.50[0.27,0.67]
Sympathy	4.08(0.62)	4.07(0.48)	4.44(0.41)	4.10	(2,28.08)	.027*	0.44[0.17,0.69]
Intelligence	3.78(0.49)	3.61(0.69)	4.04(0.55)	3.05	(2,28.60)	.063	0.37[0.13,0.58]
User Traits							
Immersive Tendency	4.63(0.63)	4.74(0.66)	4.53(0.74)	0.68	(2,29.75)	.515	0.24[0.04,0.53]
General Motivation	4.55(0.38)	4.51 (0.46)	4.43(0.61)	0.29	(2,29.66)	.749	0.24[0.07,0.52]
VR Sickness							
Pre	8.13(8.48)	9.83(7.92)	8.40(7.24)	0.47	(2,29.44)	.627	0.21[0.04,0.43]
Post	8.30(8.14)	12.08(8.02)	8.18(7.76)	2.42	(2,29.01)	.107	0.34[0.10,0.58]
Sense of Embodiment							
Body Ownership	4.17(1.27)	4.14(1.56)	4.42(1.38)	0.23	(2,29.55)	.793	0.21[0.03,0.48]
Agency	5.55(0.89)	5.45(0.98)	5.86(0.90)	1.38	(2,29.21)	.267	0.30[0.08,0.73]
Change	3.55(1.06)	3.34(1.41)	3.02(1.67)	0.80	(2,29.34)	.458	0.25[0.06,0.59]

5 Discussion

We partially accept H1: Feedback with high performance specificity resulted in greater perceived realism compared to medium specificity. Our system-driven adaptation of feedback had a positive impact on the perception of the VR experience, similar to what research on user-driven adaptations has shown [14, 42, 55]. We found no significant effects for the other subscales or between the low and medium specificity groups. Interestingly, mean values for the low specificity group were higher than those of the medium group across all subscales, surpassing the scores of the high specificity group on two subscales (performance and sound), though none of these differences reached statistical significance. While exploratory, this pattern points to a more nuanced relationship between feedback specificity and the perception of the VR experience.

Latoschik and Wienrich's Congruency and Plausibility Model [35] may help interpret our findings. The model proposes that alterations in the manipulation space, i.e., design choices, impact an overall perceived plausibility, which subsequently influences the emergence of intermediate factors like presence. Following this model, high performance specificity may have been perceived as more plausible than medium specificity, aligning better with participants' expectations of a competent coach in a realistic training scenario. In contrast, participants may have found the low specificity condition — characterized by a quieter coach who mainly explained the exercises — more coherent and less intrusive, thus more plausible than the medium specificity condition, which involved more frequent but potentially less helpful interventions.

A possible reason for the limited effects regarding all subscales except perceived realism is the subtlety of our feedback manipulation, especially between medium and high, which may not have sufficiently affected ratings across all subscales, some of which (e.g., interface quality, possibility to act, sound) are strongly influenced by core technological factors like visual fidelity and responsiveness.

In summary, our findings indicate that performance specificity of feedback influences subjective perceptions of the VR experience, particularly perceived realism. However, while the significant difference between medium and high specificity supports this hypothesis for realism, the absence of significant effects on other subscales, along with exploratory trends favoring low specificity, point to a more complex interplay. Thus, although highly performance-specific feedback may enhance the experience, minimal specificity was not clearly less effective and warrants further exploration.

We reject H2: More performance-specific feedback did not result in higher motivation. The results do not show any significant differences across the IMI subscales. We found a non-significant trend ($p = .053$) that perceived relatedness tended to be higher for participants interacting with the virtual coach providing highly performance-specific feedback, compared to those in the medium or low specificity conditions. Relatedness refers to the perceived connection with others [16, 17]. If future studies confirm this trend, performance-specific feedback might serve as a tool to positively impact the development of the relationship between the user and the coach, which is essential for effective coaching [8, 12, 31, 38].

We designed the coach's feedback based on previous research suggesting that feedback must promote athlete competence while providing an appropriate challenge to foster motivation [17, 38]. We do not consider insufficient challenge to be a likely explanation for the lack of effect on motivation. While the exercises used are standard in knee rehabilitation and on the easy side for healthy individuals, informal feedback after the experiment indicated that participants still found them challenging. This suggests that the exercises were sufficiently demanding for the study population. However, our results show that the level of performance specificity of feedback applied in this study did not significantly impact participants' perceptions of their own competence. The coach's feedback was more about how to perform the exercise well regarding movement speed and consistency. However, a successful repetition was counted regardless of how quickly or evenly a participant performed the movement as long as their movement completely covered the predefined range. This was a conscious design decision, as we wanted to prevent repeated failures in one of the conditions, which could have confounded the whole VR experience. However, it is plausible that feedback could have a stronger impact on motivation if users perceived more tangible, direct benefits from it.

Consistent with the findings for H1, we observed a similar non-significant trend, with the descriptive values for the low specificity group exceeding (or, in the case of tension, falling below, as lower values indicate less tension) those of the medium specificity group on four of the five subscales (enjoyment, perceived competence, tension, and relatedness). This is surprising as the coach in the low specificity condition did not provide any motivational or corrective feedback during the workout. However, it is in line with our assumption that participants in the medium and high conditions felt like they did not truly need the feedback for successful execution.

In summary, our findings lead to a rejection of the hypothesis that performance-specific feedback improves motivation. Nevertheless, the results highlight the complexity of motivational responses in VR exercise applications and may inform future design considerations.

We partially accept H3: Performance specificity of feedback influenced the perception of the virtual coach's anthropomorphism and sympathy. The virtual coach providing highly performance-specific feedback was perceived as more anthropomorphic than the virtual coach in both other conditions, and more sympathetic than the virtual coach offering medium-specific feedback. However, we did not find a significant difference in the perceived intelligence of the virtual coach, even though the descriptive values showed a similar trend ($p < .07$). It is possible that more pronounced and sophisticated performance-specific feedback could reveal differences that were not detectable in this study.

Consistent with the findings for H1 and H2, the differences between the low and the medium condition were not significant but we found the same interesting pattern that the medium condition received the lowest descriptive values on all three measures related to the perception of the coach. We cannot rule out the possibility that there was a confounding variable that we did not assess, such as perceived uncanniness or discomfort with the virtual coach in one of the conditions. In the medium condition, participants experienced increased interaction with the coach compared to the low condition, but the feedback may not have been sufficiently helpful or specific to justify the heightened presence. This could have drawn attention to the artificial nature of the coach, potentially leading to a subtle sense of unease or detachment. In contrast, the minimal interaction in the low specificity condition may have avoided this effect altogether, which would be in line with Baylor and Kim [5] who found that sometimes less can be more when it comes to non-verbal behavior of pedagogical agents. Meanwhile, in the high specificity condition, the more performance-specific feedback might have outweighed potential uncanny effects by providing more valuable guidance. Ratings were highest for all three measures in the high specificity condition. These results fit with our previous assumption that the coach giving the most specific feedback was perceived as the most plausible, whereas the coach with the medium-specific feedback may have been perceived as the least plausible. However, this hypothesis is valid only if plausibility correlates positively with the perceived anthropomorphism, intelligence, and sympathy of a virtual human. What makes virtual humans plausible and how plausibility influences specific intermediate factors remains an area of ongoing research [35, 40].

In summary, we partially accept our third hypothesis. The observed increase in perceived anthropomorphism and sympathy in the high specificity condition suggests that performance-specific feedback may positively influence users' basic perceptions of the virtual coach. This, in turn, may contribute to fostering a good relationship between the user and the virtual coach.

Design suggestion: Our study demonstrates that performance specificity of feedback can positively influence users' perception of both the VR experience as a whole, as well as the virtual coach in particular. We were able to produce positive effects using easily available low-cost methods to increase specificity of feedback that were only based on the completed number of repetitions and various time measurements. Therefore, we suggest incorporating specific references to the user's exercise performance, e.g., completed number of repetitions, individual times, and performance-specific corrections when providing workout-accompanying feedback through an embodied virtual coach in a VR exercise application. For performance-specific feedback to be truly effective, it should be based on sufficiently rich information about the user's performance and context within the application. Without this level of detail, the feedback may lack relevance and fail to deliver meaningful support to the user.

6 Limitations and Future Work

The following section outlines the limitations of our study and offers directions for future research. First, the level of performance specificity of feedback implemented in our study is far from fully exploiting the technological possibilities available, and the differences in the feedback between conditions were rather subtle. We were able to produce positive effects already with our low-cost method that required no additional sensors or expert-knowledge on machine learning and artificial intelligence. However, future studies could achieve greater effects by including more performance-specific data in the design of the feedback, e.g., training history, positional or rotational data, and physiological measurements. In particular, we believe that a more comprehensive utilization of technological possibilities and performance-specific feedback that is more tightly bound to the success of the user, might strengthen their perceived competence thereby affecting their motivation.

Enhancing the performance specificity of feedback will very likely also impact the perception of the coach. In our study, the descriptive values regarding the perception of the coach, in particular, for anthropomorphism and intelligence, were quite low (below the midpoint of the scale) across all conditions, suggesting considerable room for improvement of the coach's behavioral realism as well as the performance-specific feedback. While we carefully designed and tested the coach's behavior including facial and body movements, all interactions were pre-recorded, and the coach only responded to the participant's actions at predefined intervals. In the medium and high specificity condition, the coach repeated himself several times due to the limited number of available voice lines. Research indicates that behavioral realism, e.g., realistic gaze behavior, significantly impacts the perception of virtual humans [24].

In this study, we chose a virtual coach with a stereotypical athletic appearance. However, Baylor et al. [6] suggest that the appearance of a virtual coach can impact its effectiveness. Alternative designs for the virtual coach's appearance may be more suitable to enhance participant motivation and the application's overall effectiveness. Therefore, the interaction between performance-specific feedback and the virtual coach's appearance requires further investigation, particularly concerning the relationship between the athlete and the virtual coach. Our findings indicate that performance-specific feedback influences users' basic perceptions of the virtual coach. Future research should explore the impact on additional intermediate factors such as social presence, rapport, and trust, which focus not only on the perception of the virtual coach but also on the user-coach relationship.

Furthermore, the lack of significant difference between the low specificity condition and both other conditions requires further investigation. Future studies should investigate in more detail how different nuances of performance-specific feedback influence the plausibility of the VR experience in general and of the virtual coach in particular [35, 40]. The relatively broad confidence intervals observed for most non-significant measures indicate a degree of uncertainty in the effect size estimates, which is likely influenced by the sample size and data variability. Increasing the number of participants in future studies could help narrow these intervals, leading to more precise and reliable estimates of the true effects.

Lastly, our participant pool consisted of university students resulting in a relatively homogenous sample, particularly in terms of age. This limits the ability to generalize the findings to more diverse populations. This consideration is especially important when designing VR exercise applications for user groups with distinct needs, e.g., older adults or individuals undergoing rehabilitation.

7 Conclusion

In a user study, we examined the impact of performance-specific verbal feedback provided by an embodied virtual coach within a VR exercise application on participants' perception of the VR experience, motivation, and perception of the coach. While we found no impact on motivation, our findings suggest that the level of performance specificity of feedback plays a significant role in shaping both the VR experience and users' perceptions of the virtual coach. Specifically, high levels of specificity have the potential to enhance both. Therefore, we derive the design suggestion to include precise performance-specific details when designing feedback for a virtual coach in the VR exercise context. Interestingly, while not statistically significant, the low specificity condition showed higher descriptive scores than the medium specificity condition on most measures, indicating a potential trend that may warrant further investigation. It suggests that, under certain circumstances — which should be examined more closely in future work — it may be more effective to provide minimal performance-specific feedback than to offer feedback that is only mildly specific. Our study provides first insights into the influence of performance-specific feedback from an embodied coach in the context of VR exercise applications. Future research should explore the nuances of performance specificity in verbal feedback to further clarify its impact on users and inform the development of evidence-based design guidelines that connect essential design choices to intermediate factors influencing the success of a VR exercise application.

Acknowledgments

This research was supported by the Bavarian State Ministry for Digital Affairs in the project XR Hub (Grant A5-3822-2-16).

References

[1] A. Baca. Perspectives of artificial intelligence in training and exercise. In M. J. Blondin, I. Fister Jr., and P. M. Pardalos, eds., Artificial Intelligence, Optimization, and Data Sciences in Sports, pp. 265–274. Springer Nature Switzerland, 2025. doi: 10.1007/978-3-031-76047-1_9 2

[2] A. Bartl, C. Merz, D. Roth, and M. E. Latoschik. The effects of avatar and environment design on embodiment, presence, activation, and task load in a virtual reality exercise application. In IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 260–269, 2022. doi: 10.1109/ISMAR55827.2022.00041 1, 3, 4, 5

[3] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics, 1(1):71–81, 2009. doi: 10.1007/s12369-008-0001-3 5

[4] A. L. Baylor. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535):3559–3565, 2009. doi: 10.1098/rstb.2009.0148 2

[5] A. L. Baylor and S. Kim. Designing nonverbal communication for pedagogical agents: When less is more. Computers in Human Behavior, 25(2):450–457, 2009. doi: 10.1016/j.chb.2008.10.008 8

[6] A. L. Baylor and Y. Kim. Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In International conference on intelligent tutoring systems, pp. 592–603. Springer, 2004. doi: 10.1007/978-3-540-30139-4_56 8

[7] A. L. Baylor, E. Shen, D. Warren, and S. Park. Supporting learners with math anxiety: The impact of pedagogical agent emotional and motivational support. In Workshop on "Social and Emotional Intelligence in Learning Environments", Int. Conference on Intelligent Tutoring Systems, 2004. doi: 10.1007/978-3-540-30139-4_125 2

[8] R. Beale and C. Creed. Affective interaction: How emotional agents affect users. International Journal of Human-Computer Studies, 67(9):755–776, 2009. doi: 10.1016/j.ijhcs.2009.05.001 2, 7

[9] N. Benar, S. Fatahi Milasi, and A. Nazarian. The behavioral intention of coaches in virtual reality: The roles of perceived usefulness, enjoyment, performance expectancy, self-efficacy, and ease of use. Journal of New Studies in Sport Management, 5(3):1187–1198, 2024. doi: 10.22103/jnssm.2024.22372.1229 1

[10] R. Bi, D. Gao, X. Zeng, and Q. Zhu. Lazier: A virtual fitness coach based on ai technology. In 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), pp. 207–212. IEEE, 2022. doi: 10.1109/ICISCAE55891.2022.9927664 2

[11] B. Biancardi, C. Wang, M. Mancini, A. Cafaro, G. Chanel, and C. Pelachaud. A computational model for managing impressions of an embodied conversational agent in real-time. In 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–7, 2019. doi: 10.1109/ACII.2019.8925495 2

[12] I. Buchem, O. Stamm, S. Vorwerg, K. Kruschel, and K. Hildebrand. Evaluation of rapport in human-agent interactions with a vr trainer after a 6-week exergame training for senior users with hypertension. In AHFE Int. Conf. on Human Factors in Virtual Environments and Game Design, vol. 50, 2022. doi: 10.54941/ahfe1002071 2, 7

[13] S. Cacho-Elizondo, J.-D. Lázaro Álvarez, and V.-E. Garcia. Assessing the opportunities for virtual, augmented, and diminished reality in the healthcare sector. In L. Menvielle, A.-F. Audrain-Pontevia, and W. Menvielle, eds., The Digitization of Healthcare: New Challenges and Opportunities, pp. 323–344. Palgrave Macmillan UK, London, 2017. doi: 10.1057/978-1-349-95173-4_18 1

[14] Z.-H. Chen, H.-D. Lu, and C.-H. Lu. The effects of human factors on the use of avatars in game-based learning: Customization vs. non-customization. International Journal of Human–Computer Interaction, 35(4-5):384–394, 2019. doi: 10.1080/10447318.2018.1543090 2, 3, 7

[15] R. Corbett, M. Partington, L. Ryan, and E. Cope. A systematic review of coach augmented verbal feedback during practice and competition in team sports. International Journal of Sports Science & Coaching, 19(2):864–881, 2024. doi: 10.1177/17479541231218665 3

[16] E. L. Deci and R. M. Ryan. Promoting self-determined education. Scandinavian journal of educational research, 38(1):3–14, 1994. doi: 10.1080/0031383940380101 2, 3, 7

[17] E. L. Deci and R. M. Ryan. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychol. Inq., 11(4):227–268, 2000. doi: 10.1207/S15327965PLI1104_01 2, 3, 7, 8

[18] M. Dobiasch, B. Krenn, R. P. Lamberts, and A. Baca. The effects of visual feedback on performance in heart rate-and power-based-tasks during a constant load cycling test. Journal of Sports Science & Medicine, 21(1):49, 2022. doi: 10.52082/jssm.2022.49 2

[19] E. Etienne, M. Ristorcelli, S. Saufnay, A. Quilez, R. Casanova, M. Schyns, and M. Ochs. A systematic review on the socio-affective perception of ivas' multi-modal behaviour. In Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents, pp. 1–

10, 2024. doi: 10.1145/3652988.3673943 2, 3, 4

[20] C. for Self-Determination Theory. Intrinsic motivation inventory (imi). https://selfdeterminationtheory.org/intrinsic-motivation-inventory/, 2022. last accessed: 03.08.2025. 5

[21] A. für Sport und Gesundheit. Geschlechtsspezifische verteilung bei der ausbildung zum personal trainer in deutschland in den jahren 2015 bis 2017. https://de.statista.com/statistik/daten/studie/829841/umfrage/geschlechtsspezifische-verteilung-bei-der-ausbildung-zum-personal-trainer/, 2018. last accessed: 03.08.2025. 3

[22] S. Gago-Masague, T. M. Chen, and G.-P. Li. Promoting physical exercise through embodied trainers: A systematic literature review. In L. Menvielle, A.-F. Audrain-Pontevia, and W. Menvielle, eds., The Digitization of Healthcare: New Challenges and Opportunities, pp. 293–322. Palgrave Macmillan UK, London, 2017. doi: 10.1057/978-1-349-95173-4_17 1, 3

[23] D. Gall and M. E. Latoschik. Visual angle modulates affective responses to audiovisual stimuli. Computers in Human Behavior, 109:106346, 2020. doi: 10.1016/j.chb.2020.106346 2

[24] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In Proc. of the SIGCHI conference on Human factors in computing systems, pp. 529–536, 2003. doi: 10.1145/642611.642703 8

[25] J. A. García-Herrero, D. Soto-García, R. J. Carcedo, I. Martínez-Martín, and P. Delgado-Floody. Analysis of the relationship between satisfaction with the coach and the effect of comparative social feedback in elite female handball players. Int. j. of environmental research and public health, 19(13):7680, 2022. doi: 10.3390/ijerph19137680 2

[26] F. Guay, G. A. Mageau, and R. J. Vallerand. On the hierarchical structure of self-determined motivation: A test of top-down, bottom-up, reciprocal, and horizontal effects. Personality and Social Psychology Bulletin, 29(8):992–1004, 2003. doi: 10.1177/0146167203253297 5

[27] N. Hamzeheinejad, D. Roth, S. Monty, J. Breuer, A. Rodenberg, and M. E. Latoschik. The impact of implicit and explicit feedback on performance and experience during vr-supported motor rehabilitation. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pp. 382–391. IEEE, 2021. doi: 10.1109/VR50410.2021.00061 1, 3

[28] M. C. Howard. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Computers in Human Behavior, 70:317–327, 2017. doi: 10.1016/j.chb.2017.01.013 1, 2

[29] F. Hülsmann, S. Kopp, and M. Botsch. Automatic error analysis of human motor performance for interactive coaching in virtual reality. arXiv preprint, 2017. doi: 10.48550/arXiv.1709.09131 2

[30] W. A. IJsselsteijn, Y. A. W. d. Kort, J. Westerink, M. d. Jager, and R. Bonants. Virtual fitness: Stimulating exercise behavior through media technology. Presence: Teleoperators and Virtual Environments, 15(6):688–698, 12 2006. doi: 10.1162/pres.15.6.688 1, 2

[31] S. Jowett and A. Poczwardowski. Understanding the coach-athlete relationship. In S. Jowett and D. Lavallee, eds., Social Psychology in Sport, pp. 3–14. Human Kinetics, Champaign, IL, 2007. doi: 10.5040/9781492595878.ch-001 2, 7

[32] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. Presence: Teleoperators and Virtual Environments, 21(4):373–387, 2012. doi: 10.1162/PRES_a_00124 4

[33] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. Applied ergonomics, 69:66–73, 2018. doi: 10.1016/j.apergo.2017.12.016 5

[34] H. C. Lane and N. L. Schroeder. Pedagogical agents. In B. Lugrin, C. Pelachaud, and D. Traum, eds., The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, p. 307–330. Association for Computing Machinery, New York, NY, USA, 1 ed., 2022. doi: 10.1145/3563659 2, 3

[35] M. E. Latoschik and C. Wienrich. Congruence and plausibility, not presence: Pivotal conditions for xr experiences and effects, a novel approach. Frontiers in Virtual Reality; Virtual Reality and Human Behaviour; Presence and Beyond: Evaluating User Experience in AR/MR/VR, 3, 2022. doi: 10.3389/frvir.2022.694433 7, 8, 9

[36] M. Lee, J. Son, J. Kim, and B. Yoon. Individualized feedback-based virtual reality exercise improves older women's self-perceived health: A randomized controlled trial. Archives of Gerontology and Geriatrics, 61(2):154–160, 2015. doi: 10.1016/j.archger.2015.06.010 2

[37] B. Lugrin, C. Pelachaud, and D. Traum, eds. The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition, vol. 37. Association for Computing Machinery, New York, NY, USA, 1 ed., 2021. doi: 10.1145/3477322 5

[38] G. A. Mageau and R. J. Vallerand. The coach–athlete relationship: A motivational model. Journal of sports science, 21(11):883–904, 2003. doi: 0.1080/0264041031000140374 2, 3, 7, 8

[39] P. Mair and R. Wilcox. Robust statistical methods in r using the wrs2 package. Behavior research methods, 52(2):464–488, 2020. doi: 10.3758/s13428-019-01246-w 6

[40] D. Mal, E. Wolf, N. Döllinger, M. Botsch, C. Wienrich, and M. E. Latoschik. Virtual human coherence and plausibility; towards a validated scale. In 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, pp. 788–789, 2022. doi: 10.1109/VRW55335.2022.00245 8, 9

[41] A. Marougkas, C. Troussas, A. Krouska, and C. Sgouropoulou. Crafting immersive experiences: A multi-layered conceptual framework for personalized and gamified virtual reality applications in education. In K. Kabassi, P. Mylonas, and J. Caro, eds., Novel & Intelligent Digital Systems: Proceedings of the 3rd International Conference (NiDS 2023), pp. 230–241. Springer Nature Switzerland, Cham, 2023. doi: 10.1007/978-3-031-44097-7_25 2

[42] S. Pardini, S. Gabrielli, M. Dianti, C. Novara, G. M. Zucco, O. Mich, and S. Forti. The role of personalization in the user experience, preferences and engagement with virtual reality environments for relaxation. International Journal of Environmental Research and Public Health, 19(12), 2022. doi: 10.3390/ijerph19127237 2, 3, 7

[43] L. G. Pelletier, S. C. Dion, M. Slovinec-D'Angelo, and R. Reid. Why do you regulate what you eat? relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. Motivation and emotion, 28(3):245–277, 2004. doi: 10.1023/B:MOEM.0000040154.40922.14 5

[44] A. S. Praetorius, L. Krautmacher, G. Tullius, and C. Curio. User-avatar relationships in various contexts: Does context influence a users' perception and choice of an avatar? In Mensch Und Computer 2021, MuC '21, p. 275–280. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3473856.3474007 2

[45] F. Richlan, M. Weiß, P. Kastner, and J. Braid. Virtual training, real effects: a narrative review on sports performance enhancement through interventions in virtual reality. Frontiers in Psychology, 14, 2023. doi: 10.3389/fpsyg.2023.1240790 1

[46] G. Robillard, S. Bouchard, P. Renaud, and L. G. Cournoyer. Presence questionnaire (revised version). http://w3.uqo.ca/cyberpsy/wp-content/uploads/2019/04/QEP_vf.pdf, 2002. last accessed: 03.08.2025. 5

[47] D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (veq). IEEE Transactions on Visualization and Computer Graphics, 26(12):3546–3556, 2020. doi: 10.1109/TVCG.2020.3023603 4, 5

[48] D. Roth, J.-L. Lugrin, M. E. Latoschik, and S. Huber. Alpha ivbo – construction of a scale to measure the illusion of virtual body ownership. In Proc. of CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2875–2883, 2017. doi: 10.1145/3027063.3053272 5

[49] S. Schneider and F. Kummert. Comparing the effects of social robots and virtual agents on exercising motivation. In International Conference on Social Robotics, pp. 451–461. Springer, 2018. doi: 10.1007/978-3-030-05204-1_44 5

[50] E. H. Simpson and P. D. Balsam. The behavioral neuroscience of motivation: an overview of concepts, measures, and translational applications. Behavioral neuroscience of motivation, pp. 1–12, 2015. doi: 10.1007/7854_2015_402 2

[51] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton. A survey of presence

and related concepts. ACM Computing Surveys (CSUR), 50(6):1–39, 2017. doi: 10.1145/3134301 2, 5

[52] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138 2

[53] M. Slater et al. Measuring presence: A response to the witmer and singer presence questionnaire. Presence: teleoperators and virtual environments, 8(5):560–565, 1999. doi: 10.1162/105474699566477 2

[54] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments, 6(6):603–616, 1997. doi: 10.1162/pres.1997.6.6.603 2

[55] C.-I. Teng. Customization, immersion satisfaction, and online gamer loyalty. Computers in Human Behavior, 26(6):1547–1554, 2010. doi: 10.1016/j.chb.2010.05.029 2, 3, 7

[56] K. M. Tsiouris, V. D. Tsakanikas, D. Gatsios, and D. I. Fotiadis. A review of virtual coaching systems in healthcare: closing the loop with real-time feedback. Frontiers in Digital Health, 2:567502, 2020. doi: 10.3389/fdgth.2020.567502 2

[57] A. Uribe-Quevedo, B. Kapralos, D. R. Gualdron, A. Dubrowski, S. Perera, F. Alam, and S. Xu. Physical and physiological data for customizing immersive vr training. In 2021 IEEE/ACIS 20th Int. Fall Conference on Computer and Information Science (ICIS Fall), pp. 156–160, 2021. doi: 10.1109/ICISFall51598.2021.9627412 2

[58] R. J. Vallerand. Intrinsic and extrinsic motivation in sport and physical activity: A review and a look at the future. In Handbook of Sport Psychology, chap. 3, pp. 59–83. John Wiley & Sons, Ltd, 2007. doi: 10.1002/9781118270011.ch3 2, 3

[59] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. IEEE Transaction on Visualization and Computer Graphics, 24(4):1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629 2, 4

[60] S. Wei, D. Freeman, and A. Rovira. A randomised controlled test of emotional attributes of a virtual coach within a virtual reality (vr) mental health treatment. Scientific Reports, 13(1):11517, 2023. doi: 10.1038/s41598-023-38499-7 2

[61] T. G. Weimann, H. Schlieter, and A. B. Brendel. Virtual coaches: Background, theories, and future research directions. Business & Information Systems Engineering, 64(4):515–528, 2022. doi: 10.1007/s12599-022-00757-9 1

[62] P. Weiss, E. Keshner, and M. Levin. Virtual reality for physical and motor rehabilitation. Springer, 2014. doi: 10.1007/978-1-4939-0968-1 4

[63] M. D. Westby, A. Brittain, and C. L. Backman. Expert consensus on best practices for post–acute rehabilitation after total hip and knee arthroplasty: a canada and united states delphi study. Arthritis care & research, 66(3):411–423, 2014. doi: 10.1002/acr.22164 2

[64] B. G. Witmer, C. J. Jerome, and M. J. Singer. The Factor Structure of the Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 14(3):298–312, 2005. doi: 10.1162/105474605323384654 5

[65] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3):225–240, 1998. doi: 10.1162/105474698565686 5