



# Desktop virtual reality as an exposure method for test anxiety: quantitative and qualitative feasibility study

Luca Chittaro<sup>1</sup> · Marta Serafini<sup>1</sup>

Received: 24 October 2022 / Revised: 22 August 2023 / Accepted: 7 September 2023  
© The Author(s) 2023

## Abstract

Test anxiety is an emotional state characterized by subjective feelings of discomfort, fear, and worry that can considerably affect students' academic performance. Virtual Reality exposure (VRE) is a promising approach to address test anxiety, but the few VRE systems for test anxiety in the literature concern only written exams. Since oral exams elicit more anxiety than written exams, the availability of VRE systems for oral exams would be precious to a large population of students worldwide. Another limitation of existing VRE systems for test anxiety is that they require the availability of a head-mounted display, posing a barrier to widespread use. This paper aims to address both issues, proposing a VRE system that deals with oral exams and can be used with common PC displays. The design of the proposed system is organized in three oral test scenarios in which a virtual agent acts as the student's examiner. The virtual examiner behaves friendly in the first scenario and increasingly reduces its friendliness in the two subsequent scenarios. The paper assesses the feasibility for VRE of the proposed system with two complementary methods. First, we describe a quantitative user study of the three system scenarios, showing that they induce increasing levels of anxiety. Second, we present a qualitative thematic analysis of participants' post-exposure interviews that sheds further light on the aspects of the virtual experience that contributed to eliciting negative or positive affect in participants, and provides insights for improving VRE systems for test anxiety.

**Keywords** Virtual reality · Test anxiety · Virtual agent · Oral exams

## 1 Introduction

Test anxiety is an emotional state characterized by subjective feelings of discomfort, fear, and worry about the evaluation made by an external authority [1]. Anxiety can have a considerable negative impact on students' academic performance [2, 3], hence motivating the study of techniques that could help students in reducing it.

---

✉ Marta Serafini  
marta.serafini@uniud.it

<sup>1</sup> Human-Computer Interaction Lab (HCI Lab), University of Udine, via delle Scienze 206, 33100 Udine, Italy

In vivo and imaginal exposure approaches have traditionally been used in psychology to treat various types of anxiety. Both approaches expose the individual to his or her feared situations over time, by either using the real situation (in vivo) or asking the individual to imagine it (imaginal). Repeated exposure to the feared situation causes a desensitization that gradually lowers the elicited anxiety, making the situation less fearful. Over the past two decades, Virtual Reality Exposure (VRE) has been increasingly used as an alternative to in vivo and imaginal exposure in the treatment of anxiety [4–7]. VRE involves repeatedly exposing people to the feared stimulus in a virtual environment (VE), i.e. a simulated environment created through computer-generated images. VRE systems for test anxiety should thus reproduce different exam situations [8]. However, the simulated exams should be able to produce anxiety responses in the student to support the desensitization process.

Unfortunately, the literature on VRE systems for test anxiety is very limited, with only three papers available on this specific topic [8–10]. Moreover, they suffer from three major limitations. First, not all of them assess the capability of their system to produce anxiety responses, which is a pre-requisite for possible VRE use. Second, all of them concern a written test. Since oral tests elicit higher levels of test anxiety than written tests [11], the availability of VRE systems for oral tests would be precious to a large population of students worldwide. Third, existing VRE systems for test anxiety require the availability of a head-mounted display, posing a barrier to widespread use, especially at home. This paper aims to address all three limitations by proposing a VRE system for test anxiety that focuses on oral tests reproduced on common computer displays, and by assessing with a study the feasibility of the system for VRE. In the proposed system, a virtual examiner<sup>1</sup> (VX) conducts the oral exam. The VX can perform three different sets of behavior, creating different oral exam scenarios aimed at eliciting three levels of increasing anxiety for VRE purposes.

This paper presents the following contributions: (1) it proposes the first VRE system for test anxiety that deals with oral exams; (2) it conducts the first feasibility study of a VRE system for oral exams, showing its capability to induce different levels of anxiety; (3) while VRE systems for test anxiety were based on immersive VR so far, the system proposed and studied in this paper employs desktop VR; (4) unlike other studies of VRE systems for test anxiety, the study also exploited a visual annotation tool, proposed and described in this paper, that allows to identify the instants in which participants remember they had felt positive or negative affect. The information collected through the tool is also useful in guiding further development of the system, as described in the paper.

## 2 Related work

The literature has proposed several VRE systems [12–24] aimed at fear of public speaking that allow the user to give a speech to a virtual audience. Studies [18–22, 24] have shown that some of these systems are able to increase the level of anxiety and distress in participants, confirming feasibility for VRE in public speaking. Other studies [13–17, 23] have used the systems for treatment, showing that the VRE approach is successful in reducing the level of public speaking anxiety. However, such scenarios are different from

---

<sup>1</sup> We will refer to the virtual examiner (VX) with the pronoun “it” for the rest of the paper.

**Table 1** Comparison of the main differences of the present paper with studies conducted on VRE systems for test anxiety

Reference	Type of exam	Type of display	Studied VR effect	Participants
[8]	Written	Head-mounted display	Elicited anxiety	21 university students
[9]	Written	Head-mounted display	Elicited anxiety	22 adolescents
[10]	Written	Head-mounted display	NA	NA
This paper	Oral	Desktop monitor	Elicited anxiety	36 university students

those needed in test anxiety, which should instead reproduce the situations experienced by students in school exams.

To the best of our knowledge, there are only three proposals of VRE systems for test anxiety in the literature [8–10]. Table 1 synthetically highlights the main differences between those works and the present paper.

TAVE (Test Anxiety Virtual Environments) [8] is an immersive VRE system that provides users with three VEs, to be experienced in chronological order: student's home, representing the day before and then the morning of the exam; followed by subway journey to the exam location; and finally school corridor and classroom where the exam takes place. A feasibility study was conducted to assess whether the different VEs of TAVE could elicit significantly different responses in students. The system was tested on 11 students with high levels of test anxiety and ten students with low levels of test anxiety. After experiencing each VE, students' level of anxiety was measured with the State Anxiety Inventory and Subjective Units of Discomfort Scale questionnaires. Results showed that students with high test anxiety experienced higher levels of anxiety than students with low test anxiety in all the VEs. However, the level of anxiety in the two groups did not progressively increase with the sequence of VEs but peaked during the subway journey.

Kwon et al. [9] proposed an immersive VRE system that includes two VEs: house on the day before the exam and school on the day of the exam. They assessed the feasibility of the VRE system in inducing different levels of test anxiety in adolescents. However, in addition to experiencing the VEs in an exposure session, participants performed also a meditation session to regulate anxiety immediately after the exposure session. The system was tested on 21 adolescents whose general anxiety and test anxiety were measured with two self-report questionnaires at the beginning of the study, before they were exposed to the first VE. The study also measured participants' heart rate variability with a physiological sensor during VEs exposure. After each session, adolescents' level of anxiety was measured with the visual analogue scale. The main results of the study showed a significant difference in anxiety only between exposure and meditation sessions. The study does not instead specify if there were significant differences in perceived anxiety between the exposure sessions.

Luo et al. [10] developed a VRE system for test anxiety containing three VEs aimed at eliciting increasing anxiety levels: home on the day before the exam, school entrance on the day of the exam, and classroom where the exam takes place. The system was assessed on middle school adolescents: unfortunately, while the paper states that participants experienced varying degrees of anxiety during VE exposures, it does not illustrate a study that might provide evidence about the system's ability to elicit increasing levels of anxiety.

The VRE systems for test anxiety described above have three aspects in common.

First, they concern a written test. In TAVE [8], students are seated in a classroom and take a written test with multiple-choice, general knowledge questions. In [9], adolescents are

seated in a classroom where a written test is about to start, but the test itself is not simulated. In [10], students are seated in a classroom and, after selecting their preferred subject from a list of five options, proceed to take a written test with questions from a college entrance test. In some educational systems, students take oral tests as an adjunct or an alternative to written tests, and several studies have highlighted the benefits of oral tests as assessment method [25–27]. The reliance on oral tests changes from country to country, for example they are central to the Italian educational system [28, 29], while the US educational system prefers written tests [26, 27, 30, 31]. As already mentioned, oral tests elicit higher levels of anxiety than written tests [11], and the availability of VRE systems for oral tests could thus benefit a large population of students worldwide. Moreover, the need for engaging in interactions with the professor in oral exams introduces an additional challenge for students who also suffer from social anxiety. Therefore, it is crucial to provide students with scenarios in which they are not only exposed to an exam situation but also to the interaction with the professor. For these reasons, we focused our proposed system on oral exams.

Second, current VRE systems for test anxiety require using a head-mounted display. Although this can increase users' sense of immersion, it restricts the use of the system to laboratory settings and to people who own a head-mounted display. Conversely, the use of widely available hardware would allow a much larger number of students to benefit from the VRE system, also at home. Moreover, although modern head-mounted displays have made significant progress in limiting symptoms of motion sickness, there is still the risk of negative side effects, such as nausea, headaches, and dizziness [32, 33]. On the other hand, the use of a VRE system using commonly available hardware would help minimize health risks. Furthermore, existing studies on effects of immersive VR on users show that VR can lead to more visual fatigue than traditional screens [34]. For these reasons, the VRE system we propose is meant for common computer monitors.

Third, the studies on VRE systems for test anxiety have exposed participants to the different VEs in a fixed order. In [8], both the group with high test anxiety and the group with low test anxiety were exposed to the three VEs in chronological order: first the home, then the subway, then the classroom VE. In [9], all participants were exposed to the two VEs in the same order: first the house, then the school VE. Exposing all participants to the VEs in a fixed order could lead to order effects, i.e., the participants' responses to the VEs could be affected by the participants' responses elicited in the previously experienced VEs. On the contrary, to prevent order effects, the order of conditions was counterbalanced in our study, i.e., all six possible orders of conditions were created, and then different participants were assigned a different order as shown in Fig. 1.

## 3 Materials and methods

### 3.1 Participants

The study was approved by the Institutional Review Board of the University of Udine. A sample of 37 (32 male, 5 female) participants was recruited for the study. Participants were volunteers who received no compensation. Their age ranged between 20 and 35 ( $M=22.28$ ,  $SD=2.78$ ), and they were recruited through direct contact among undergraduate Computer Science students of the University of Udine. Since the proposed VRE system concerns the simulation of an oral test, we looked for participants who were likely to take an oral test in the immediate future as they could be more representative of the intended users of the

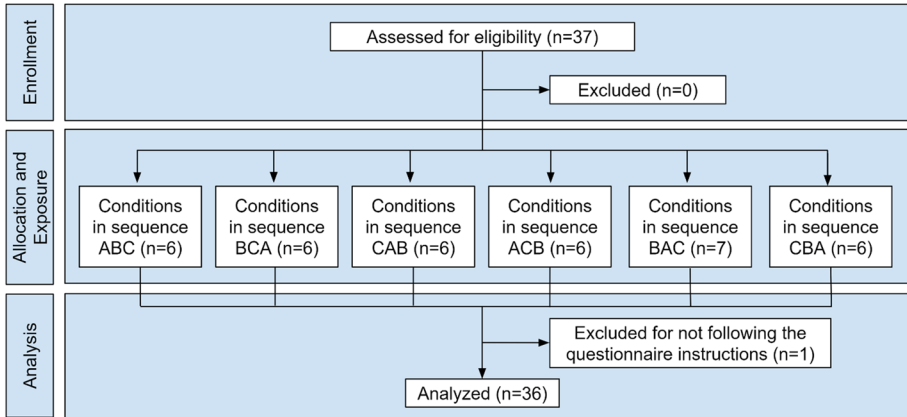


Fig. 1 Participant flow diagram

system. For this reason, students who had completed all the exams in their curriculum were not considered eligible. One participant was excluded from the analysis because he did not follow the instructions in completing some of the questionnaires. Figure 1 shows the participant flow diagram.

Although the recruited sample was predominantly male, it must be noted that there is substantial evidence from several studies that females tend to report greater anxiety [35] and higher intensity of emotional experiences [36, 37] than males. We thus reasoned that, if male predominance in our sample was possibly going to influence the results, the influence would have been in the direction of obtaining a smaller, not a greater, response in the intensity of anxiety.

### 3.2 The proposed VE

The system was developed in Unity version 2019.4.18f1. As a 3D model of the VX, we used the “Business\_Female\_03” character from Microsoft’s Rocketbox library [38]. The VX spoke with the “Elsa (Neural)” voice of the Azure Cognitive Services text-to-speech, with pitchDelta parameter set to -12.

In the VE, the VX sat behind a desk in an office. We defined three sets of VX behaviors (A, B, C), intended to elicit three increasing levels of anxiety. The three sets aimed at making the VX appear respectively friendly (set A), only partially friendly (set B),



Fig. 2 Examples of different VX behaviors: nodding to the student in set A (a); one of the distracted positions in set B (b); head scratching in set C (c)

**Table 2** Description of behaviors in A, B, C. The micro-movements mentioned in the IDLE BEHAVIOR row were obtained through head motion capture of a human actor

Behavior	A	B	C
<b>IDLE BEHAVIOR</b>	Smiling facial expression with micro-movements of the head and eyes, and eye-blink	Neutral facial expression with micro-movements of the head and eyes, and eye-blink	Annoyed facial expression with micro-movements of the head and eyes, and eye-blink
<b>BEHAVIOR 1</b>	Smiling and nodding (slowly moving the head up and down three times)	Slowly moving the head from right to left and left to right three times while looking at the student with a sad expression	Looking at the student with an annoyed expression, changing the head tilt five times
<b>BEHAVIOR 2</b>	Tilting the head to the left, smiling	Looking distracted, staring at a spot toward the upper left corner of the room	Looking up with a reflective expression, then slowly swinging the head with an annoyed expression
<b>BEHAVIOR 3</b>	Looking at the student, smiling and changing head position and orientation five times	Looking at the student, changing head position and orientation five times	Looking at the student with an annoyed expression while rubbing the hands on the thighs
<b>BEHAVIOR 4</b>	Crossing the arms by placing one hand on the opposite hip and the other hand on the opposite shoulder, looking at the student and smiling	Raising the forearms, bringing the hands to shoulders height on the two sides of the body, without changing the height of the elbows, with a doubtful expression	Scratching the head with both hands, with an annoyed expression
<b>BEHAVIOR 5</b>	Tilting the head to the right, smiling	Looking at the student, changing the head tilt five times	Texting on a smartphone kept on the knees under the desk, lowering the head to look at and smile at the phone
<b>BEHAVIOR 6</b>	Looking at the student, changing head tilt five times, smiling	Looking distracted, turning the head to the right, then pausing, then turning the head further to the right	Checking the time on the wristwatch with an annoyed expression

and unfriendly (set C). Figure 2 shows an example of a behavior from each set, and Table 2 describes all behaviors. Behaviors were chosen based on the indications about their likely effect on users, available in the literature on non-verbal communication [39–43] and virtual agents [44–46].

In set A, the VX maintained a smiling expression to convey agreement [42] and a positive attitude [43–45] (IDLE BEHAVIOR in Table 2). The VX performed all other behaviors in set A with a more pronounced smile than IDLE BEHAVIOR to make the perception of the VX's facial expression more recognizable, as recommended in [45]. In addition, to further represent agreement and a positive attitude, set A included nodding [42–45] (BEHAVIOR 1) and tilting the head [39] (BEHAVIOR 2 and BEHAVIOR 5). Placing one hand on the opposite hip and the other hand on the opposite shoulder [46] (BEHAVIOR 4) were introduced to represent interest.

In set B, the VX kept a neutral expression on its face, not showing any signs of positive or negative attitude (IDLE BEHAVIOR in Table 2). Behaviors performed by the VX in set B to convey a neutral or slightly negative attitude, boredom, or disengagement included shaking its head [44, 45] (BEHAVIOR 1 in Table 2), turning its head away from participant [39, 44] (BEHAVIOR 2 and BEHAVIOR 6), raising its eyebrows [44] while raising its arms (BEHAVIOR 4).

In set C, the VX maintained a frowning expression to convey a negative attitude [45] (IDLE BEHAVIOR in Table 2). The VX performed all other behaviors in set C with a more pronounced frowning face than IDLE BEHAVIOR to make the perception of the VX's facial expression more recognizable, as recommended in [45]. Behaviors performed by the VX in set C to convey a negative attitude or discomfort included performing a frowning and thoughtful expression and then shaking the head [45] (BEHAVIOR 2 in Table 2), scratching the head (BEHAVIOR 4) [47], rubbing the hands on the thighs (BEHAVIOR 3) [47], facing down [45] to use a smartphone (BEHAVIOR 5), checking the time on the wristwatch [43] (BEHAVIOR 6).

Each set also included neutral behaviors (i.e., BEHAVIOR 3 and BEHAVIOR 6 in set A; BEHAVIOR 3 and BEHAVIOR 5 in set B; BEHAVIOR 1 in set C). In these behaviors, the VX changed head position and orientation several times while maintaining a smiling (A), neutral (B) or frowning (C) expression.

The VE was displayed from the viewpoint of a student sitting in front of the VX. The VX first greeted the student, then asked him or her three test questions, and finally informed him or her that the exam was over.

The proposed system offers the possibility of customizing the questions (users can define a pool of questions from which the VX randomly selects its questions), but we chose not to use this feature in the study in order to ask exactly the same questions to each participant. In choosing the subject of the questions, we reasoned that different familiarity of participants with the topic could affect the level of anxiety they could experience. To prevent this confounding factor, the questions in the study concerned a topic (Basics of International Law) that was unrelated to the participants' degree curriculum. In addition, we checked that participants were not possibly familiar with the topic for other reasons. Each question was followed by 30 seconds of silence, during which the VX performed three behaviors from the assigned set of behaviors, following the order illustrated in Table 3. The timing of events was the same in the three conditions, and is described in detail by Table 3.

**Table 3** Script of the oral exam experience. The VX spoke in Italian, all sentences have been translated here into English for reader's convenience

Sequence of events	Behaviors performed by the VX
Opening greeting	The VX says: <i>"Good morning. Let's start with the first question"</i>
First question	The VX says: <i>"Describe to me the general principles that are universally recognized by civilized nations"</i>
Wait	The VX performs the following sequence of behaviors from the assigned set (A, B or C): [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 1] for 10 seconds [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 2] for 10 seconds [IDLE BEHAVIOR] for 3 seconds
Introduction to the next question	The VX says: <i>"Let's proceed with another topic"</i>
Second question	The VX says: <i>"What are the conditions for diplomatic protection?"</i>
Wait	The VX performs the following sequence of behaviors from the assigned set (A, B or C): [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 3] for 10 seconds [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 4] for 10 seconds [IDLE BEHAVIOR] for 3 seconds
Introduction to the next question	The VX says: <i>"Let's proceed with another question"</i>
Third question	The VX says: <i>"Tell me about the European Convention on Human Rights"</i>
Wait	The VX performs the following sequence of behaviors from the assigned set (A, B or C): [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 5] for 10 seconds [IDLE BEHAVIOR] for 5 seconds [BEHAVIOR 6] for 10 seconds [IDLE BEHAVIOR] for 3 seconds
Ending greeting	The VX says: <i>"That was the last question. It can be enough. Goodbye"</i>

### 3.3 Visual Annotation Tool (VAT)

One of the goals of this study was to better understand which aspects of the experience elicit affective responses in participants. To do so, we wanted to identify the exact time instants that had elicited an affective response in participants during exposure, distinguishing between positive responses (i.e., time instants in which participants had felt at ease) and negative responses (i.e., time instants in which participants had felt distressed). Given the central role of the VX in the experience, we were particularly interested in identifying which VX-related factors elicited such responses in participants. To support the achievement of these goals, we developed a Visual Annotation Tool (VAT) as a software-based measuring instrument. The VAT allows one to replay the exposure experience, and to mark any time instant as positive or negative by pressing two different keys on the keyboard, made easier to recognize by a red sticker and a green sticker. Once the replay of the experience is complete, the VAT displays the marked time instants on a timeline, representing positive and negative instants as green and red



dots, respectively. Then, each time instant can be selected to replay the corresponding part of the experience and categorize them in a list of pre-defined categories.

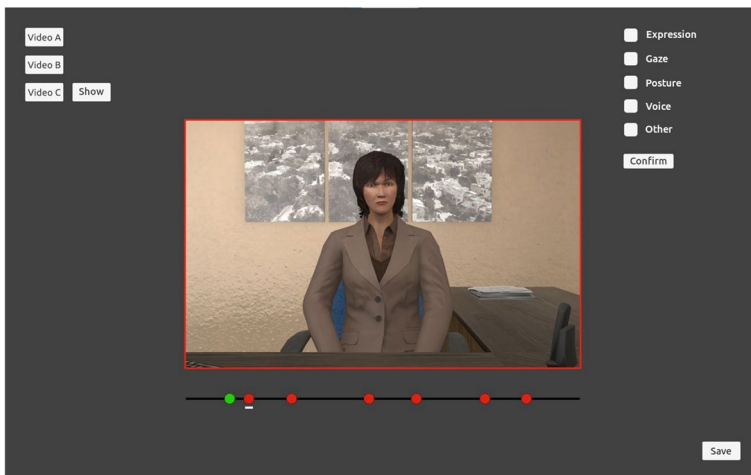
In this study, we used the VAT as follows. After experiencing the VRE system, participants used the VAT to replay the three conditions in the same order they had experienced. Participants were instructed to mark the time instants that made them feel at ease or distressed during the first exposure by pressing the green or red key on the keyboard, respectively. Once the replay of all conditions was complete, the VAT displayed green and red dots on the timeline for each condition, allowing the experimenter to identify the specific instants during which the participants respectively felt at ease or distressed. Figure 3 shows the timeline of condition C with the positive and negative dots identified by one of the participants.

Then, the experimenter used the three timelines to conduct an interview with participants to investigate the factors that had made them feel distressed or at ease during exposure. To do so, we used as selectable categories in the VAT three VX factors (i.e., facial expression, gaze, and posture), and an additional item “other” that allowed for the inclusion of free descriptive text. The experimenter selected each marked time instant to replay the corresponding part of the experience and asked the participant to indicate which factors had caused the elicited feeling. Based on participant’s answer, the experimenter categorized the time instant in the VAT. In Fig. 3, the experimenter is examining the replay of the second time instant from the left in the timeline (the VAT highlights the selected time instant with a white line below it).

### 3.4 Measures and hypotheses

#### 3.4.1 Self-reported social interaction anxiety

We administered the Italian adaptation of the Social Interaction Anxiety Scale (SIAS) [48] to measure the level of anxiety among participants in general social interactions. The SIAS



**Fig. 3** The annotation tool, as seen by the experimenter. In this screenshot, there is only one time instant in which the participant felt at ease (represented by the first dot from the left in the timeline data visualization, colored green) while all other dots (colored red) identify time instants in which the participant felt distressed. The dot selected by the experimenter is highlighted by a white line. The items listed in the upper right part of the figure allow the experimenter to assign one or more categories to the selected dot after interviewing the participant about that time instant

**Table 4** VX attitude questionnaire items

1	Is it angry, tense, nervous?
2	Is it serene, relaxed, calm?
3	Is it rude, distracted?
4	Is it sociable, kind?

is a 19-item self-report questionnaire which describes anxious reactions that can occur during social interactions [49]. For each item, respondents rate how true the statement is for them on a 5-point Likert-type scale (0=“not at all”, 4=“extremely”). Total score range from 0 to 76. Higher scores indicate higher social interaction anxiety. Consistently with the literature on VRE systems for social anxiety [50–56], we chose to use the SIAS as an instrument to measure participant’s trait anxiety in the specific context of social interaction.

### 3.4.2 Self-reported anxiety

We used the Visual Analogue Scale for Anxiety (VAS-A) as an instrument to assess participants’ state anxiety [57, 58] during the exposure to each condition. The scale was a 10 cm long line, with “not at all anxious” and “very anxious” printed at its left and right ends, respectively. Participants reported their score by drawing a vertical mark on the scale.

### 3.4.3 VX attitude

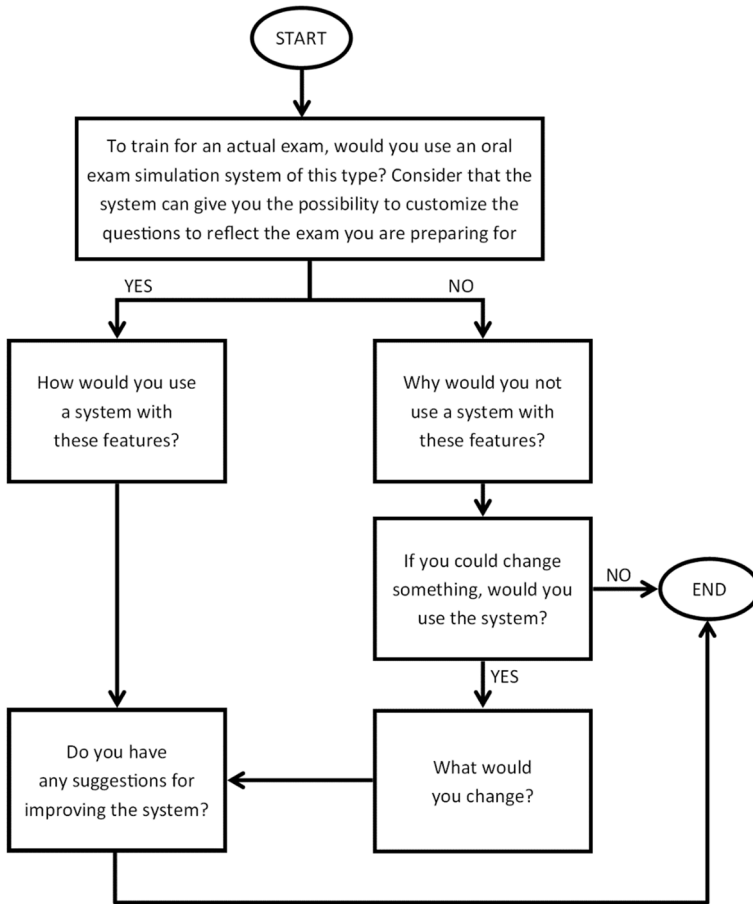
We administered a 4-item questionnaire (Table 4) that asked participants to rate aspects of the VX’s attitude they perceived on a 7-point Likert-type scale (1=“not at all”, 7=“very”). To calculate the score, the scale of the second and fourth items was inverted, and the answers were averaged. Higher scores indicate a more negative attitude. Cronbach’s alpha in the three conditions was respectively 0.70 (A), 0.60 (B), 0.86 (C). We performed an exploratory factor analysis with principal component extraction and Oblimin rotation to evaluate factorial validity. Bartlett’s test was significant ( $p < .001$  in all conditions) and KMO was greater than .60 in set A (0.70) and set C (0.75), while it was 0.54 in set B. The analysis confirmed the intended one-factor structure that explained respectively 57.27% of variance in set A, 48.64% in set B, 72.63% in set C.

### 3.4.4 Elicited positive and negative responses

We measured the number of times participants reported a feeling of ease (positive responses) or distress (negative responses) with the VAT. For each condition, the two counts were used as an indication of positive and negative responses elicited.

### 3.4.5 Factors associated to positive and negative responses

As described in Section 3.3, factors associated to positive and negative responses were collected through the participant’s interview and stored in the VAT. The aim was to



**Fig. 4** Flow diagram of the structured qualitative interview. The interview was conducted in Italian, all sentences have been translated here into English for reader's convenience

identify the VX factors (and other possible factors) that had a greater impact in eliciting responses in participants.

### 3.4.6 Qualitative interview

After conducting the first part of the interview with the VAT as described in Section 3.3, we further interviewed participants to gather comments about how to improve the system. The experimenter asked a sequence of structured open-ended questions, following the diagram in Fig. 4. The data collected were analyzed by employing thematic analysis.

### 3.4.7 Hypotheses

We formulated the following hypotheses:

- H1. The three conditions would produce three different, increasing values of anxiety because the three sets of behavior of the VX aim at making the VX appear friendly in condition A, only partially friendly in condition B, and unfriendly in condition C. We expect that the increasingly less friendly behavior of the VX elicit a correspondingly increasingly higher level of anxiety in participants.
- H2. The three conditions would produce three different, increasingly negative perceptions of VX attitude because the VX has been designed to behave friendly when it performs set A, and to increasingly reduce its friendliness when it performs sets B and C.
- H3. The three conditions would produce three different, decreasing (resp. increasing) counts of positive (resp. negative) responses elicited because the VX's behavior is more friendly when it performs set A, less friendly when it performs set B, and unfriendly when it performs set C. We thus expect that sets A, B, and C elicit a progressively decreasing (resp. increasing) number of positive (resp. negative) responses in participants.

### 3.5 Procedure

We aimed at maintaining study design consistency with previous studies that assessed the feasibility of VRE systems for social anxiety [52, 55], fear of public speaking [21, 22] and test anxiety [8, 9]. Participants were tested individually in a 50-minute session that started with filling the SIAS. To prevent the spill over of anxiety elicited by a condition to the next condition, each condition was preceded by a two-minute period during which participants sat in a comfortable position while listening to calm music and watching a series of relaxing images of natural scenarios (such as forests, pools and hills) through a 24-inches desktop monitor positioned in front of them. Participants were asked to imagine being the student in the oral test, and told that they did not need to answer the questions audibly. After exposure to each condition (A, B, C) via the desktop monitor, participants filled the VAS-A and VX perception questionnaires. The same sequence (relax, exposure, questionnaire) was followed for each condition. Participants were exposed to the three conditions in counterbalanced order as illustrated in detail by Fig. 1. After experiencing all the conditions, participants replayed the conditions in the same order they were presented previously, using the previously described VAT to mark the time instants of the experience in which they remembered to have felt at ease or distressed. Finally, participants were interviewed: for each time instant marked by the participant with the VAT, the experimenter watched the corresponding part of the experience in the VAT together with the participant, and discussed with him or her the reasons why that experience had caused a positive or negative reaction. Then, the experimenter categorized the participant's answer in the VAT. Finally, the experimenter asked a sequence of structured open-ended questions, following the diagram in Fig. 4.

## 4 Quantitative results

All analyses were conducted using SPSS version 28.0.1.0. A repeated-measure ANOVA was used to compare the effects of the VX's behaviors (A, B, C) on VAS-A, VX attitude, and counts of elicited positive and negative responses. If Mauchly's test indicated a

**Table 5** ANOVA results

	A	B	C	Main effect		
	Mean (SD)	Mean (SD)	Mean (SD)	F	p	$\eta_p^2$
VAS-A	1.86 (1.65)	3.09 (1.75)	4.28 (2.28)	F(2, 70)=23.70	<0.001	0.82
Negative attitude of VX	2.00 (0.77)	4.60 (1.09)	5.58 (1.13)	F(1.7, 59.51)=147.47	<0.001	0.97
Elicited positive responses	3.94 (1.80)	0.56 (0.97)	0.06 (0.23)	F(1.35,47.09)=112.39	<0.001	0.84
Elicited negative responses	0.67 (1.04)	2.69 (1.77)	5.11 (2.30)	F(2, 70)=85.73	<0.001	0.82

violated assumption of sphericity, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. ANOVA results are shown in Table 5.

ANOVA revealed a main effect of state anxiety measured with VAS-A, and Bonferroni post hoc comparison found a significant difference for all pairs of conditions (A vs. B,  $p<0.01$ ; B vs. C,  $p<0.01$ ; A vs. C,  $p<0.001$ ).

ANOVA revealed a main effect of VX's attitude, and Bonferroni post hoc comparison found a significant difference for all pairs of conditions (A vs. B,  $p<0.001$ ; A vs. C,  $p<0.001$ ; B vs. C,  $p=0.001$ ).

ANOVA revealed a main effect for both counts of elicited positive and negative responses. Bonferroni post hoc comparison found a significant difference for all pairs, both with positive responses (A vs. B,  $p<0.001$ ; A vs. C,  $p<0.001$ ; B vs. C,  $p<0.05$ ) and negative responses (all three pairs,  $p<0.001$ ).

To explore if participants with different levels of social anxiety were affected differently by the proposed system, for each dependent variable we contrasted the group of participants with an above-median score ( $n=18$ ) with the group of remaining participants ( $n=18$ ). The median score in the SIAS was 30.50 ( $M=30.64$ ,  $SD=12.20$ , range 10-60). A 3x2 mixed design ANOVA (between-subjects: above-median vs. below-median SIAS scores, within-subjects: A vs. B vs. C) revealed no significant differences between the two groups for any measure, and no interaction between the two variables.

## 5 Qualitative results

To conduct the qualitative analysis of the participants' interviews, we initially transcribed the audio recordings verbatim. Then, following Braun and Clarke's method [59], we performed a thematic analysis on the transcripts, identifying and categorizing common and prominent themes. The analysis included the following:

1. Reading and re-reading transcripts to familiarize with data;
2. Coding features of interest in the dataset and collating relevant data for each code;
3. Grouping all codes into potential themes, collecting all pertinent data for each theme, and organizing themes into levels (i.e., major themes or sub-themes within them). When a subject was particularly large or complex, it was split into one or more sub-themes;
4. Determining the significance of the themes and sub-themes with respect to the coded extracts and the whole dataset;
5. Refining each theme and sub-theme, generating clear definitions and names.

The corresponding author performed the steps described above and coded the data. However, since the process of defining the codes and applying them to the dataset can be biased by subjective interpretation [60], the validity and reliability of the themes must be confirmed through additional coding. Following [61], the data were coded also by an independent external coder, who was not involved in our research, and used a codebook we provided. The codebook listed the themes and sub-themes identified by the thematic analysis. For each code, it provided a label and a complete description with inclusion and exclusion criteria. We also explained to the external coder that he could use multiple codes on the same text fragment. Both coders used Taguette [62], an open-source web-based CAQDAS (Computer Assisted Qualitative Data Analysis Software), to code the data. SPSS version 28.0.1.0 was used to compute Cohen's Kappa to assess inter-rater reliability [63, 64]. The overall kappa coefficient was 0.80, showing a strong level of agreement [65].

Themes fell into three topic areas:

- *Suggestions*: themes capturing suggestions for enhancing the system;
- *System*: themes related to system features;
- *Voice*: themes related to the VX's voice.

In the following, we summarize the themes and sub-themes of the three topic areas. More details can be found in Appendix 1.

Tables 6, 7, and 8 summarize the themes and sub-themes belonging to the topic areas of *Suggestions*, *System*, and *Voice*, respectively.

## 6 Discussion

### 6.1 Elicited anxiety

Results confirmed the feasibility of a desktop VR system for potential use in VRE for test anxiety. Indeed, they confirmed our hypothesis on anxiety (H1): conditions A, B, and C elicited different, increasing anxiety levels in participants. The ability of the VRE system to elicit different levels of anxiety extends the findings of other studies of VRE systems for test anxiety [8, 9]. Those systems simulated written exams, used immersive VR, and exposed participants to different conditions in a fixed order. In contrast, our study focused on the simulation of oral exams using desktop VR and exposed participants to different conditions in a counterbalanced order to prevent order effects. Moreover, no significant differences were found between participants with above-median SIAS scores and the remaining participants. This result supports the feasibility of the proposed system in eliciting different levels of anxiety regardless of participants' trait social anxiety.

### 6.2 Perceived VX's attitude

Results confirmed our hypothesis on perceived VX's attitude (H2): the three conditions produced increasing negative perceptions. The ability to elicit increasing levels of anxiety by changing the attitude displayed by the VX extends the results of the literature. Studies of VRE systems for fear of public speaking [20, 22] showed that exposure to a group or crowd of agents with different attitudes can elicit different levels of anxiety in participants. Our study showed that similar results can be obtained with a single virtual agent.

**Table 6** Themes and sub-themes of the Suggestions topic area

Theme	Sub-theme	Description	Participants
Virtual examiner (VX)	Animations	Some participants suggested to increase the level of realism of the VX's animations	P4, P17, P20, P27, P35
	Artificial Intelligence	Some participants suggested to add artificial intelligence to the VX to make it able to recognize students' correct and wrong answers	P5, P18, P30, P35
	Customization	Some participants suggested to introduce the possibility of customizing the gender, appearance, or behavior of the VX	P14, P18, P22, P27
	Verbal requests	Some participants suggested to make the verbal requests made by the VX more varied, such as "explain in more detail", "make an example", ...	P2, P6, P12, P13, P23
	Physical appearance	Some participants suggested to increase the level of realism of the VX's physical appearance	P1, P5, P17, P29, P34
System	Set of behaviors	A few participants suggested to increase the variety of behaviors and expressions that the VX can perform	P1, P8, P18
	Voice	Some participants suggested to use a more realistic voice, varying how the VX uses it in the three levels of difficulty	P2, P8, P9, P18, P28
	Level of difficulty	A few participants suggested to add the ability to manually choose the level of difficulty on which to practice the oral exam	P18, P22
	Recording	A few participants suggested to add the ability to record the oral exam session	P11, P30
Virtual environment	Answer time	Some participants suggested additional functions concerning the time available to answer questions	P11, P18, P24
	Immersion	A few participants were interested in trying an immersive version of the system with a headset	P5, P8
	Audience	A few participants suggested to add the presence of an audience to the virtual environment	P8, P25, P29
	Realism	A few participants suggested to increase the level of realism of the virtual room	P5, P25

**Table 7** Themes and sub-themes of the System topic area

Theme	Sub-theme	Description	Participants
Critiques	Low emotional involvement	Some participants felt that the system only partially evoked the emotions of a real oral test	P1, P6, P13, P17, P23, P34
	Unuseful as a study method	A few participants said that the system would be unuseful for them as a study method, mentioning they are used to a different method for preparing exams	P3, P7
Purpose	Emotional preparation	Many participants would use the system to train themselves in managing the emotional aspects of oral exams	P4, P8, P9, P10, P12, P14, P15, P18, P19, P21, P24, P26, P27, P28, P31, P32, P33, P35
	Study method	Some participants would use the system as a study method to prepare for the exam	P2, P4, P6, P22, P29, P30
Intention to use	If improved	A few participants stated that they would use the system if it is improved	P1, P7, P17
	No	A few participants stated they would not use the system in any case	P3, P36
	Yes	Many participants stated that they would use the system	P2, P4, P8, P9, P10, P11, P12, P15, P16, P18, P19, P20, P21, P22, P24, P25, P26, P27, P28, P29, P30, P31, P32, P33, P35



**Table 8** Themes and sub-themes of the Voice topic area

Theme	Sub-theme	Description	Participants
Sound	Flat	A few participants described the VX's voice as flat, without inflections	P4, P9, P28
	Neutral	A few participants described the VX's voice as neutral	P12, P22
	Robotic	Some participants described the VX's voice as robotic, mechanical	P9, P25, P31, P36
	Strong	Some participants stated that the VX's voice was different between (at least two of the three) conditions	P13, P20, P31, P32, P35
Change perception	None	Some participants stated that VX's voice was the same in the three conditions	P2, P8, P15, P21, P24

A previous study showed that the attitudes of a virtual job interviewer, controlled by the experimenter with a Wizard-of-Oz technique (i.e., the user perceives a direct interaction with the virtual job interviewer, while in fact the experimenter decides and commands the responses of the virtual interviewer in real-time), can elicit different levels of anxiety [52]. Our study showed how similar effects can be obtained in a virtual oral exam administered by an agent that is controlled by sets of predefined behaviors.

### 6.3 Participants' positive and negative responses

Unlike previous virtual agent and VRE studies, we created a visual annotation tool (VAT) to identify the time instants in which participants felt positive or negative affect. The VAT also supported an interview to collect factors that contributed to participants' positive or negative responses. It is worth noting that information collected through the VAT is also useful for further guiding the development of the system because it helps in identifying the sources of participants' ease or distress. By knowing the different aspects of the VE that cause more at ease or distress in participants, it is possible to further improve the VRE system by inserting the right stressors in the different scenarios of the system. For these reasons, its use could be considered for other VRE studies. The VAT allowed us to obtain information that would not have been collected using the questionnaires alone. Results collected through the VAT confirmed our hypothesis on elicited positive and negative responses (H3): the VX's behaviors produced three different, decreasing counts of positive responses elicited and three different, increasing counts of negative responses elicited. Counts of positive and negative responses elicited in participants were consistent with the increasing anxiety elicited by the three conditions (A, B, C), and were also consistent with the increasing negative perception of VX's attitude. Results obtained by using the VAT in interviewing participants indicated expression and posture of the VX as particularly influential factors, consistently with previous studies that highlighted the relevance of virtual agent posture and facial expressions in eliciting users' emotions [40]. Several participants remembered they felt a negative response when the VX suspended eye contact with them, consistently with studies that highlighted the importance of eye contact in positive interpersonal interaction [41, 42].

Additional participants' comments revealed that positive responses were often elicited when the VX appeared interested and listening, or calm and relaxed. On the contrary, negative responses were often elicited when the VX appeared distracted, uninterested, judgmental or impatient. Some participants mentioned that their negative responses originated at times from a negative perception of their performance during the simulation rather than from specific VX's behavior.

In summary, participants' feedback indicated that the facial expressions, posture, eye contact, level of interest, and general attitude of the VX played a central role in influencing participants' emotional responses.

### 6.4 Participants' feedback

In the final interview, the system received a high level of consensus: 25 participants stated that they would use the system; a few others would use it after some improvement (n=3). Twenty-three participants would use it to train for the oral exam (as a study method, n=6; or to learn how to manage emotional aspects, n=18). The fact that most participants would use the system, and the current unavailability of VRE systems for oral exams, motivate the

importance of continuing the research, and of further development of the system. The thematic analysis described in Section 5 and Appendix 1 allowed to identify several comments and suggestions that can be useful to inform the design of VRE systems for test anxiety. It is interesting to note that the aspect of the system on which participants reflected more frequently was the VX's voice. Five participants suggested to make changes to the VX's voice to better reflect the VX's attitude. They suggested improving the VX's voice to make it more realistic, less flat, and more expressive. In particular, one participant suggested introducing variations in the tone of the VX's voice so that when the VX does not put him at ease, its tone of voice also reflects this discomfort. Studies of virtual agents have shown that the emotions they convey through facial expressions, head movements, and voice elicit a greater response in the user while the agent is speaking rather than listening [66]. For this reason, changing VX's voice to reflect VX's attitude may enhance anxiety elicitation in participants. It is worth noting that 5 participants believed the VX's tone of voice changed between conditions, while it actually did not. Changes they perceived went in the direction of consistency with perceived VX's attitude. This can be an interesting aspect for further research because it suggests that VX behavior and facial expressions may cause illusory voice changes that are consistent with participants' expectations derived from behavioral cues.

During the interview, participants identified the three conditions as three different levels of difficulty, where set A was the easiest level, set B was the intermediate level while set C was the most difficult level. The order of difficulty perceived by participants was consistent with the increasingly negative perception of the VX's attitude and the level of anxiety increasingly elicited.

During the interview, most participants' suggestions to improve the system were focused on increasing level of realism in different aspects, i.e., quality or variety of animations, appearance of VX or VE, voice or range of expressions of the VX. In particular, four participants suggested that the VX should be customizable not only in terms of behavior but also in appearance, in order to closely resemble the real professor they feared. Allowing students to personalize the appearance of the VX is an aspect that should be considered for future improvements of VRE systems focused on oral exams because a previous study showed that individuals who conducted an oral presentation before virtual agents with an appearance similar to real people known to them experienced higher anxiety than when the presentation was conducted in front of unfamiliar virtual agents [67]. Likewise, a VX resembling the student's real professor in appearance could potentially elicit heightened levels of anxiety, similar to what he or she might experience during an actual oral exam.

## 6.5 Limitations and future research

To the best of our knowledge, our paper is the first proposal and feasibility study of a VRE system for test anxiety that deals with oral exams. While existing VRE systems for test anxiety simulate predefined written exams whose questions cannot be customized, our system allows users to customize questions and can thus better adapt to students' needs by offering an experience closer to the exam they are preparing for. However, there are some limitations that should be taken into account. First, the study was conducted on a predominantly male sample. As mentioned earlier (Section 3.1), the possible consequences of a predominant male sample on anxiety measures have been examined in the literature: several studies show that females tend to report greater anxiety [35] and higher intensities of emotional experience [36, 37] than males. Moreover, females exhibit higher levels of test anxiety than males [68, 69]. Therefore, if male predominance in our sample influenced the results, it

might have attenuated the intensity of anxiety. Replicating the experiment with a gender-balanced sample might thus produce higher anxiety values and provide a more accurate representation of anxiety experienced across genders. Second, all participants in the study shared a common academic background in Computer Science, therefore they all possessed a similar set of skills, academic experiences, and perspectives related to this discipline. This common background might not have fully captured the variability and nuances that could have emerged from participants with different academic backgrounds. As a result, the conclusions of this study might be limited in their generalizability to other fields. Third, we asked participants to use the VAT to mark the time instants of the experience in which they remembered feeling at ease or distressed during the first exposure. This may introduce a potential recall bias. Participants might have been asked to mark their affective responses during the first exposure, but we ruled out this possibility to maintain their focus on the task and enable them to identify more closely with the role of the student in an oral exam. In this way, they should be more likely to feel emotions similar to those they would experience during a real oral exam. Furthermore, we opted to introduce the VAT only after participants had experienced all three conditions to prevent participants from focusing during exposure on thinking about which emotions they should mark later with the VAT, and thus letting their attention remain primarily on the task. Fourth, the evaluation of the system focused on assessing its feasibility in eliciting anxiety in students, but the system was not used in an intervention. Therefore, our next research step will be to carry out a study of the system in a home program for reducing test anxiety using a gender-balanced sample of participants. We plan to conduct a study in which participants will have the ability to define a pool of questions from which the virtual examiner will randomly select its questions. Furthermore, during the oral exam simulation, participants will be able to answer audibly to the questions while the virtual examiner will randomly perform the behaviors of the selected set. Finally, participants will be able to self-assess their performance to track their progress over time. The study will also explore the efficacy of the VRE system in reducing anxiety in participants by comparing their anxiety before and after the treatment period.

## 7 Conclusions

In this paper, we proposed a VRE system for test anxiety that deals with oral exams. In the three scenarios offered by the system, a virtual examiner conducts the oral exam, performing behaviors from one of three predefined sets, which differ in how friendly the examiner behaves. To the best of our knowledge, this is the first feasibility study of a VRE system for test anxiety that deals with oral exams. Results of the quantitative study showed that the three scenarios of the proposed system are able to elicit three different levels of increasing anxiety. The thematic analysis of the interview conducted with the participants provided further insights into the aspects of the system that contributed to eliciting positive or negative responses in them, which can help in improving the design of VRE systems for test anxiety. While this paper showed the feasibility of the proposed system for VRE, our future work will focus on using and evaluating it as a treatment for test anxiety.

## Appendix 1

In this appendix, we illustrate in more depth the results of the qualitative analysis by including sample interview extracts. In the extracts, the parts in square brackets are words or phrases added to clarify the sentence, while the parts in round brackets are additional questions asked during the interview.

### 1.1. Suggestions

#### 1.1.1. Virtual examiner (VX)

Twenty-one participants offered opinions on the virtual examiner (VX). Five of them suggested to improve the realism of the VX's physical appearance, while five suggested to improve the realism of the animations, for example:

*Make the character's attitude slightly more human. [...] I mean, make it more fluid, less robotic. – P20*

Four participants suggested to expand the VX with artificial intelligence to make it able to understand and possibly correct students' answers, for example:

*[I would like] it to react to my answers [...]. If I give a wrong answer or hesitate, the virtual teacher evaluates me, so I realize that I have to go back to study. – P5*

Four participants suggested to make the VX customizable in terms of appearance or behavior, for example:

*Maybe being able to customize the appearance of the teacher so [...] if, for example, I have in mind a professor that causes me stress [...], being able to give this [virtual] teacher the appearance of the professor I am afraid of... could help me to overcome my anxiety, make me find a way to succeed in passing the exam. – P18*

Three participants suggested to expand the range of actions and emotions that the VX can exhibit, while five others suggested to expand the diversity of its verbal requests to the user, for example:

*Make it a little more engaging, for example by maybe adding questions, I mean, not just questions and answers, but adding questions or remarks between the student's answers. (For example?) Such as, if the student answers in a certain way, the professor might ask him "Give me some examples". – P23*

Five participants suggested to improve the VX's synthetic voice using a more realistic one, or vary it among the three sets of behavior, for example:

*There should be a little bit of change in the tone of voice in which the teacher asks the questions. Because [...] in the first test, she was putting [me] a lot at ease and using a certain voice. In the second and third, she did not put people at ease but she used the same voice – P8*

#### 1.1.2. System

Seven participants offered suggestions about the VRE system. Two of them suggested to allow users to choose the level of difficulty, while two others suggested to include an option to record the session, for example:

*If it directly recorded me speaking [...], but my face too, because a lot of times it is also important to be able to communicate what you are saying, it might help me to hear how I speak, how I discuss a topic. – P11*

Two participants recommended to provide an immersive mode to use the system with a headset, while three others suggested ways to control the amount of time allowed to answer the VX's questions, for example.:

*I did not know how much time I had to answer. Maybe putting a countdown would be a source of stress because you seem to have even less time than before... but putting it to at least know approximately the duration... [or] if I know that I will receive three questions from the virtual teacher, I can more or less measure the time without having the countdown as a source of stress – P18*

### 1.1.3. Virtual environment

Four participants offered suggestions about the virtual environment. Two of them suggested to increase realism, while three others suggested to include an audience, for example:

*Seeing that there is an audience of other students behind [you], watching and listening, makes it all more anxiety-provoking – P8*

## 1.2. System

### 1.2.1 Critiques

Eight participants were dissatisfied with some aspects of the system. Six of them said it could not adequately replicate the emotions experienced during a real oral test. Two of them believed that this was due to the fact that they did not have to answer the questions during the experience, for example:

*I did not feel much engaged because I was not actually answering the questions. I mean, I had to watch the VX, but actually the level of anxiety that you feel in a real test when you have to give the answers, and maybe you are not prepared and the professor reacts like that... it is a whole different thing. – P6*

Two participants reasoned on the possible role of the system as a study method and did consider it far from their ideal study method, for example:

*I tend to study very quietly. Talking in front of a person while I study would not allow me to memorize the things I am studying. So it would not be for me. I have another method. – P7*

One participant stated that he would not use the system because he was afraid that it would provoke too many negative feelings, for example:

*I am afraid that certain negative feelings that [the system] can bring me might make my studying worse [...]. – P3*

### 1.2.2.Purpose

Twenty-three participants said they would use the system and explained why. Six of them would use it as a study technique in preparing for the oral exam, for example:

*I would use it more as a training tool to practice and so on, because it is a tool that can be used to review, to prepare for an exam. (So the way you would use the system would be as training for the exam?) Yes exactly, not so much to prepare emotionally, but more to prepare [about] the things that you should know. – P2*

Eighteen participants would use the system to prepare for the different emotional aspects of the exam, for example:

*I would use it mainly to train myself to be more relaxed during the test, because I have noticed that when I am more relaxed, let us say, I can answer with more ease and not get overwhelmed by various emotions during an oral test. – P10*

Three participants would only use the system at the highest difficulty level, for example:

*I would try to use it in the worst-case scenario... so with those facial expressions that provoked me the most anxiety, to try to maintain control. – P9*

One participant mentioned that he would utilize the system to train himself to pay greater attention to the examiner during the oral test, for example:

*[...] it can also be useful in case you are not used to looking at the professor during the oral test, in order to develop more attention towards looking at the professor, which is something that is not always obvious. – P12*

Two participants would use the system to psychologically prepare for receiving difficult questions, for example:

*Maybe I would put in [the system] the topics [...] that I feel less confident with [...] and I would try them at various difficulties: maybe when she [the examiner] is in a good mood [or] when she is irritated because the previous student did not know anything ... I would try various situations. – P31*

### 1.2.3.Intention to use

Thirty participants stated that they would (or would not) generally use the system. Two of them would not use it at all, while other three would use it only after some improvement. Twenty-five participants said they would use the system. Four of them, in particular, they would be willing to use the current system while waiting for its more advanced version, for example:

*I would say yes [I would use it]. However, [while] waiting for the developers to introduce new features, I would gladly try it out in the meantime because it still helps to feel heard [...]. It is still a useful thing. – P30*

Four participants who would use the system stated they would only use it occasionally, for example:

*I think [I would use it] based on the exam or [based on][...] how bad I feel the professor might be or how much accuracy the professor might require. – P31*

### 1.3. Voice

#### 1.3.1. Sound

Voice characteristics were discussed by eight participants. Three of them described the voice as flat and without inflection, for example:

*I was unable to feel any particular anxiety or particular tranquility from the tone of the voice. It sounded a little flat to me. – P9*

The VX's voice was described as neutral by two participants and as robotic and mechanical by four participants, for example:

*While the movements [of the virtual examiner] are quite close to reality, the voice is very robotic so it is difficult to empathize with the situation. – P9*

#### 1.3.2. Change perception

Ten participants described how the VX's voice sounded to them. Five of them felt that the voice changed with the difficulty level: two participants heard a different voice in each of the three levels, three participants in two levels, for example:

*Well, in the [two] tests in which the teacher's voice showed hurry... she evoked more anxiety so [...] I mean, in that moment I would have frozen, I would not have been able to continue the exam because I would be in trouble. But in the positive one, she was much more relaxed, yes – P32*

Five participants did not report differences in voice between the three levels of difficulty, for example:

*I found the voice to be somewhat homogeneous in all three tests. And the only thing that gave me a sense of difference between levels was the facial expressions, movements, and posture. – P15*

**Funding** Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

**Data availability** The datasets generated during or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflicts of interest** The authors have no financial or non-financial conflicts of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not



permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Irwin GS (1980) Introduction to the study of test anxiety. Test-anxiety: Theory, research and application. Lawrence Erlbaum Assoc Incorporated, Hillsdale, NJ, pp 57–78
2. Hembree R (1988) Correlates, Causes, Effects, and Treatment of Test Anxiety. *Rev Educ Res* 58:47–77. <https://doi.org/10.3102/00346543058001047>
3. Sarason IG (1963) Test anxiety and intellectual performance. *J Abnorm Soc Psychol* 66:73–75. <https://doi.org/10.1037/h0047059>
4. Carl E, Stein AT, Levihn-Coon A et al (2019) Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *J Anxiety Disord* 61:27–36. <https://doi.org/10.1016/j.janxdis.2018.08.003>
5. Fernández-Álvarez J, Di Lernia D, Riva G (2020) Virtual Reality for Anxiety Disorders: Rethinking a Field in Expansion. In: Kim Y-K (ed) *Anxiety Disorders: Rethinking and Understanding Recent Discoveries*. Springer, Singapore, pp 389–414
6. Parsons TD, Rizzo AA (2008) Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *J Behav Ther Exp Psychiatry* 39:250–261. <https://doi.org/10.1016/j.jbtep.2007.07.007>
7. Powers MB, Emmelkamp PMG (2008) Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J Anxiety Disord* 22:561–569. <https://doi.org/10.1016/j.janxdis.2007.04.006>
8. Alsina-Jurnet I, Carvallo-Beciu C, Gutiérrez-Maldonado J (2007) Validity of virtual reality as a method of exposure in the treatment of test anxiety. *Behav Res Methods* 39:844–851. <https://doi.org/10.3758/BF03192977>
9. Kwon JH, Hong N, Kim KK et al (2020) Feasibility of a virtual reality program in managing test anxiety: A pilot study. *Cyberpsychol Behav Soc Netw* 23:715–720. <https://doi.org/10.1089/cyber.2019.0651>
10. Luo D, Deng X, Luo Y, Wang G (2019) Design and implementation of virtual examination system based on unity 3D. Proceedings of the 2019 international conference on artificial intelligence and advanced manufacturing. Association for Computing Machinery, New York, pp 1–8
11. Sparfeldt JR, Rost DH, Baumeister UM, Christ O (2013) Test anxiety in written and oral examinations. *Learn Individ Differ* 24:198–203. <https://doi.org/10.1016/j.lindif.2012.12.010>
12. Alsaffar MJ (2021) Virtual reality software as preparation tools for oral presentations: Perceptions from the classroom. *Theor Pract Lang Stud* 11:1146–1160. <https://doi.org/10.17507/tpls.1110.02>
13. Harris SR, Kemmerling RL, North MM (2002) Brief Virtual Reality Therapy for Public Speaking Anxiety. *CyberPsychol Behav* 5:543–550. <https://doi.org/10.1089/109493102321018187>
14. Hinojo-Lucena F-J, Aznar-Diaz I, Cáceres-Reche M-P et al (2020) Virtual reality treatment for public speaking anxiety in students. Advancements and results in personalized medicine. *J Personal Med* 10:14. <https://doi.org/10.3390/jpm10010014>
15. Kahlon S, Lindner P, Nordgreen T (2019) Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. *Child Adolesc Psychiatry Mental Health* 13:47. <https://doi.org/10.1186/s13034-019-0307-y>
16. Lindner P, Miloff A, Fagnäs S et al (2019) Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial. *J Anxiety Disord* 61:45–54. <https://doi.org/10.1016/j.janxdis.2018.07.003>
17. Lindner P, Dagöo J, Hamilton W et al (2021) Virtual Reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial. *Cogn Behav Ther* 50:67–87. <https://doi.org/10.1080/16506073.2020.1795240>
18. Owens ME, Beidel DC (2015) Can Virtual Reality Effectively Elicit Distress Associated with Social Anxiety Disorder? *J Psychopathol Behav Assess* 37:296–305. <https://doi.org/10.1007/s10862-014-9454-x>
19. Parrish D, Oxhandler H, Duron J et al (2016) Feasibility of Virtual Reality Environments for Adolescent Social Anxiety Disorder. *Res Soc Work Pract* 26:825–835. <https://doi.org/10.1177/1049731514568897>
20. Pertaub D-P, Slater M, Barker C (2001) An experiment on fear of public speaking in virtual reality. *Stud Health Technol Inform* 81:372–378. <https://doi.org/10.3233/978-1-60750-925-7-372>

21. Pertaub D-P, Slater M, Barker C (2002) An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence: Teleoper Virtual Environ* 11:68–78. <https://doi.org/10.1162/105474602317343668>
22. Slater M, Pertaub D-P, Steed A (1999) Public speaking in virtual reality: facing an audience of avatars. *IEEE Comput Graph Appl* 19:6–9. <https://doi.org/10.1109/38.749116>
23. Sültler RE, Ketelaar PE, Lange W-G (2022) SpeakApp-Kids! Virtual reality training to reduce fear of public speaking in children – A proof of concept. *Comput Educ* 178:104384. <https://doi.org/10.1016/j.compedu.2021.104384>
24. Takac M, Collett J, Blom KJ et al (2019) Public speaking anxiety decreases within repeated virtual reality training sessions. *PLOS ONE* 14:e0216288. <https://doi.org/10.1371/journal.pone.0216288>
25. Burke-Smalley L (2014) Using Oral Exams to Assess Communication Skills in Business Courses. *Bus Prof Commun Q* 77:266–280. <https://doi.org/10.1177/2329490614537873>
26. Gharibyan H (2005) Assessing students' knowledge: oral exams vs. written tests. *SIGCSE Bull* 37:143–147. <https://doi.org/10.1145/1151954.1067487>
27. Huxham M, Campbell F, Westwood J (2012) Oral versus written assessments: a test of student performance and attitudes. *Assess Eval Higher Educ* 37:125–136. <https://doi.org/10.1080/02602938.2010.515012>
28. Pastore S, Pentassuglia M (2016) Teachers' and students' conceptions of assessment within the Italian higher education system. *Pract Res Higher Educ* 10:109–120
29. Rosalind I (2013) Italian education: Between reform and restoration. *Education and Europe: The politics of austerity*. Radicaled Books, London, UK, pp 55–85
30. Ehrlich R (2007) Giving bonus points based on oral exams. *Am J Physics* 75:374–376. <https://doi.org/10.1119/1.2431182>
31. Murillo-Zamorano LR, Montanero M (2018) Oral presentations in higher education: a comparison of the impact of peer and teacher feedback. *Assess Eval Higher Educ* 43:138–150. <https://doi.org/10.1080/02602938.2017.1303032>
32. Caserman P, Garcia-Agundez A, Gámez Zerban A, Göbel S (2021) Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality* 25:1153–1170. <https://doi.org/10.1007/s10055-021-00513-6>
33. Saredakis D, Szpak A, Birkhead B et al (2020) Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Front Human Neurosci* 14:96
34. Souchet AD, Lourdeaux D, Pagani A, Rebenitsch L (2023) A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virtual Reality* 27:19–50. <https://doi.org/10.1007/s10055-022-00672-0>
35. McLean CP, Anderson ER (2009) Brave men and timid women? A review of the gender differences in fear and anxiety. *Clin Psychol Rev* 29:496–505. <https://doi.org/10.1016/j.cpr.2009.05.003>
36. Grossman M, Wood W (1993) Sex differences in intensity of emotional experience: A social role interpretation. *J Personal Soc Psychol* 65:1010–1022. <https://doi.org/10.1037/0022-3514.65.5.1010>
37. Hess U, Senécal S, Kirouac G et al (2000) Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cogn Emot* 14:609–642. <https://doi.org/10.1080/02699930050117648>
38. Gonzalez-Franco M, Ofek E, Pan Y et al (2020) The rocketbox library and the utility of freely available rigged avatars. *Front Virtual Real* 1:561558. <https://doi.org/10.3389/frvir.2020.561558>
39. Bull P (1987) *Posture and gesture*, 1st edn. Pergamon Press, Oxford
40. Kleinsmith A, Bianchi-Berthouze N (2013) Affective Body Expression Perception and Recognition: A Survey. *IEEE Trans Affect Comput* 4:15–33. <https://doi.org/10.1109/T-AFFC.2012.16>
41. Mehrabian A (1968) Relationship of attitude to seated posture, orientation, and distance. *J Personal Soc Psychol* 10:26–30. <https://doi.org/10.1037/h0026384>
42. Tickle-Degnen L, Rosenthal R (1990) The Nature of Rapport and Its Nonverbal Correlates. *Psychol Inq* 1:285–293. [https://doi.org/10.1207/s15327965pli0104\\_1](https://doi.org/10.1207/s15327965pli0104_1)
43. Veljaca K-A, Rapee RM (1998) Detection of negative and positive audience behaviours by socially anxious subjects. *Behav Res Ther* 36:311–321. [https://doi.org/10.1016/S0005-7967\(98\)00016-3](https://doi.org/10.1016/S0005-7967(98)00016-3)
44. Chollet M, Scherer S (2017) Perception of Virtual Audiences. *IEEE Comput Graph Appl* 37:50–59. <https://doi.org/10.1109/MCG.2017.3271465>
45. Glémarec Y, Lugrin J-L, Bossier A-G et al (2021) Indifferent or Enthusiastic? virtual audiences animation and perception in virtual reality. *Front Virtual Real* 2:666232. <https://doi.org/10.3389/frvir.2021.666232>
46. Kang N, Brinkman W-P, Birna van Riemsdijk M, Neerinx M (2016) The design of virtual audiences: Noticeable and recognizable behavioral styles. *Comput Human Behav* 55:680–694. <https://doi.org/10.1016/j.chb.2015.10.008>
47. Rosenfeld HM (1966) Instrumental affiliative functions of facial and gestural expressions. *J Personal Soc Psychol* 4:65–72. <https://doi.org/10.1037/h0023514>

48. Sica C, Musoni I, Chiri LR, Bisi B, Lolli V, Sighinolfi C (2007) Social phobia scale (SPS) and social interaction anxiety scale (SIAS): Italian translation and adaptation. *B Psicol Appl* 252:59–71
49. Mattick RP, Clarke JC (1998) Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther* 36:455–470. [https://doi.org/10.1016/S0005-7967\(97\)10031-6](https://doi.org/10.1016/S0005-7967(97)10031-6)
50. Bouchard S, Dumoulin S, Robillard G et al (2017) Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *Br J Psychiatr* 210:276–283. <https://doi.org/10.1192/bjp.bp.116.184234>
51. Felnhofer A, Hlavacs H, Beutl L et al (2019) Physical Presence, Social Presence, and Anxiety in Participants with Social Anxiety Disorder during Virtual Cue Exposure. *Cyberpsychol Behav Soc Netw* 22:46–50. <https://doi.org/10.1089/cyber.2018.0221>
52. Hartanto D, Kampmann IL, Morina N et al (2014) Controlling social stress in virtual reality environments. *PLoS ONE* 9:e92804. <https://doi.org/10.1371/journal.pone.0092804>
53. Kim HE, Hong Y-J, Kim M-K et al (2017) Effectiveness of self-training using the mobile-based virtual reality program in patients with social anxiety disorder. *Comput Human Behav* 73:614–619. <https://doi.org/10.1016/j.chb.2017.04.017>
54. Kim H-J, Lee S, Jung D, et al (2020) Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: Longitudinal questionnaire study. *J Med Internet Res* 22:. <https://doi.org/10.2196/23024>
55. Morina N, Brinkman W-P, Hartanto D, Emmelkamp PMG (2014) Sense of presence and anxiety during virtual social interactions between a human and virtual humans. *PeerJ* 2014:e337. <https://doi.org/10.7717/peerj.337>
56. Morina N, Brinkman W-P, Hartanto D et al (2015) Social interactions in virtual reality exposure therapy: A proof-of-concept pilot study. *Technol Health Care* 23:581–589. <https://doi.org/10.3233/THC-151014>
57. Abend R, Dan O, Maoz K et al (2014) Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. *J Behav Ther Exper Psychiatry* 45:447–453. <https://doi.org/10.1016/j.jbtep.2014.06.004>
58. Davey HM, Barratt AL, Butow PN, Deeks JJ (2007) A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *J Clin Epidemiol* 60:356–360. <https://doi.org/10.1016/j.jclinepi.2006.07.015>
59. Braun V, Clarke V (2006) Using thematic analysis in psychology. *Qual Res Psychol* 3:77–101. <https://doi.org/10.1191/1478088706qp063oa>
60. Guest G, MacQueen KM, Namey EE (2012) *Applied Thematic Analysis*. SAGE Publications. <https://doi.org/10.4135/9781483384436>
61. DeCuir-Gunby JT, Marshall PL, McCulloch AW (2011) Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods* 23:136–155. <https://doi.org/10.1177/1525822X10388468>
62. Rampin R, Rampin V (2021) Taguette: open-source qualitative data analysis. *JOSS* 6:3522. <https://doi.org/10.21105/joss.03522>
63. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46. <https://doi.org/10.1177/001316446002000104>
64. Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 70:213–220. <https://doi.org/10.1037/h0026256>
65. McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochemia Medica* 22:276–282
66. Qu C, Brinkman W-P, Ling Y et al (2014) Conversations with a virtual human: Synthetic emotions and human responses. *Comput Hum Behav* 34:58–68. <https://doi.org/10.1016/j.chb.2014.01.033>
67. Barrett A, Pack A, Monteiro D, Liang H-N (2023) Exploring the influence of audience familiarity on speaker anxiety and performance in virtual reality and real-life presentation contexts. *Behav Inform Technol*. <https://doi.org/10.1080/0144929X.2023.2186145>
68. Núñez-Peña MI, Suárez-Pellicioni M, Bono R (2016) Gender Differences in Test Anxiety and Their Impact on Higher Education Students' Academic Achievement. *Proc Soc Behav Sci* 228:154–160. <https://doi.org/10.1016/j.sbspro.2016.07.023>
69. Bandalos DL, Yates K, Thorndike-Christ T (1995) Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *J Educ Psychol* 87:611–623. <https://doi.org/10.1037/0022-0663.87.4.611>