

Randomized Clinical Trial

Tree: Reducing the use of restrictive practices on psychiatric wards through virtual reality immersive technology training

Peter Phiri, Laura Pemberton, Yang Liu, Xiaojie Yang, Joe Salmon, Isabel Boulter, Sana Sajid, Jackie Clarke, Andy McMillan, Jian Qing Shi, Gayathri Delanerolle

Specialty type: Psychology

Provenance and peer review:

Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's classification

Scientific Quality: Grade B, Grade C

Novelty: Grade A, Grade B

Creativity or Innovation: Grade A, Grade B

Scientific Significance: Grade B, Grade B

P-Reviewer: Cheng J; Ye LS

Received: March 6, 2024

Revised: July 11, 2024

Accepted: August 20, 2024

Published online: October 19, 2024

Processing time: 224 Days and 17.3 Hours



Peter Phiri, Laura Pemberton, Joe Salmon, Sana Sajid, Gayathri Delanerolle, Department of Research & Innovation, Southern Health National Health Service Foundation Trust, Southampton SO30 3JB, United Kingdom

Yang Liu, Jian Qing Shi, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, China

Xiaojie Yang, School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, Yunnan Province, China

Isabel Boulter, Virti Healthcare, Keynsham, Bristol BS31 1SN, United Kingdom

Jackie Clarke, Neuromodulation Centre, Southern Health National Health Service Foundation Trust, Southampton SO14 0YG, United Kingdom

Andy McMillan, National Health Service Talking Therapies, Southern Health National Health Service Foundation Trust, Southampton SO50 9FH, United Kingdom

Corresponding author: Peter Phiri, BSc, PhD, RN, Academic Research, Director, Research Fellow, Department of Research & Innovation, Southern Health National Health Service Foundation Trust, Botley Road, Southampton SO30 3JB, United Kingdom.

peter.phiri@southernhealth.nhs.uk

Abstract**BACKGROUND**

Restrictive practices (RPs) are defined by measures linked to physical and chemical restraints to reduce the movement or control behaviours during any emergency. Seclusion is an equal part of RPs intended to isolate and reduce the sensory stimulation to safeguard the patient and those within the vicinity. Using interventions by way of virtual reality (VR) could assist with reducing the need for RPs as it could help reduce anxiety or agitation by way of placing users into realistic and immersive environments. This could also aid staff to and change current RPs.

AIM

To assess the feasibility and effectiveness of using a VR platform to provide reduction in RP training.

METHODS

A randomised controlled feasibility study, accompanied by evaluations at 1 month and 6 months, was conducted within inpatient psychiatric wards at Southern Health National Health Service Foundation Trust, United Kingdom. Virti VR scenarios were used on VR headsets to provide training on reducing RPs in 3 inpatient psychiatric wards. Outcome measures included general self-efficacy scale, generalised anxiety disorder assessment 7 (GAD-7), Burnout Assessment Tool 12, the Everyday Discrimination Scale, and the Compassionate Engagement and Action Scale.

RESULTS

Findings revealed statistically significant differences between the VR and treatment as usual groups, in the Everyday Discrimination Scale items Q8 and Q9: $P = 0.023$ and $P = 0.040$ respectively, indicating higher levels of perceived discrimination in the VR group. There were no significant differences between groups in terms of general self-efficacy, generalised anxiety disorder assessment 9, and Burnout Assessment Tool 12 scores. A significant difference was observed within the VR group for compassionate engagement from others ($P = 0.005$) over time. Most respondents recorded System Usability Scale scores above 70, with an average score of 71.79. There was a significant reduction in rates of RPs in the VR group *vs* treatment as usual group with a fluctuating variability observed in the VR group likely due to external factors not captured in the study.

CONCLUSION

Ongoing advancement of VR technology enables the possibility of creating scenarios and simulations tailored to healthcare environments that empower staff by providing more comprehensive and effective training for handling situations.

Key Words: Virtual reality; Restrictive practices; Inpatient wards; Restraint; Isolation; Rapid tranquilisation; Covert medication; Procedural restrictions; Health professions training

©The Author(s) 2024. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: This study explores the feasibility and effectiveness of a virtual reality (VR) platform for reducing restrictive practices in psychiatric care. Conducted at Southern Health National Health Service Foundation Trust, United Kingdom, the study used VR scenarios for training staff in three inpatient psychiatric wards. Various outcome measures like the general self-efficacy scale, generalised anxiety disorder assessment 7, and others were used. Results indicated high statistical significance for some variables, although some showed lower statistical power. The VR platform, evaluated using the System Usability Scale, was found to be highly usable. No significant differences in confidence levels were observed between the VR and treatment as usual groups, indicating the potential of VR as an effective tool for training in reducing restrictive practices.

Citation: Phiri P, Pemberton L, Liu Y, Yang X, Salmon J, Boulter I, Sajid S, Clarke J, McMillan A, Shi JQ, Delanerolle G. Tree: Reducing the use of restrictive practices on psychiatric wards through virtual reality immersive technology training. *World J Psychiatry* 2024; 14(10): 1521-1537

URL: <https://www.wjgnet.com/2220-3206/full/v14/i10/1521.htm>

DOI: <https://dx.doi.org/10.5498/wjp.v14.i10.1521>

INTRODUCTION

Restrictive practice (RP) can be defined as “any intervention used in health and care settings that restricts a person from doing something they wish to do or coerces them into doing something they do not wish to do. In the UK, this includes, but is not limited to, immobilization techniques such as physical, mechanical, and chemical restraints (sedation), seclusion, continuous observation, and restricted leave arrangements” [1]. RPs are considered controversial and potentially dangerous as they can have detrimental physical and psychological effects both on patients and staff [2,3]. Previous research has shown adverse consequences such as skin abrasions due to force, high levels of anxiety or anger, feelings of powerlessness or deprivation, or even death in the worst-case scenarios [1,4-6]. Patients and families have also reported RPs as traumatising or re-traumatising when there is a history of previous trauma [2,4,7,8]. Staff on wards have described experiences of emotional discomfort, including feelings of fear, distress, shame, and worrying about patients’ rights or physical injuries. These factors can contribute to burnout or poor retention rates of staff [1,5,6]. The mental health charity MIND has raised concerns about the inexistence of national standards or accredited training for healthcare staff in the use of RPs [2].

The National Institute for Health and Care Excellence encourages the use of the least restrictive option available [9,10]. The National Institute for Health and Care Excellence guidance for short-term management of aggression and violence in mental health outlines that “health and social care provider organisations should train staff who work in services in which

restrictive interventions may be used in psychosocial methods to avoid or minimise restrictive interventions.” [9]. Research has shown that allowing staff to develop and practice new skills to manage and de-escalate difficult situations can have a significant impact on reducing the use of RPs [3,11]. Most RPs are considered acceptable as a last resort to prevent harm but remain a controversial topic among the general public [3,8,12]. Despite increased regulations nationally and worldwide, RPs are still widely used in the National Health Service (NHS). Recent reports published by NHS digital [13] show that RPs are still common, with more than 58000 incidents in Acute Mental Health and Learning Disability settings: Around 5000 per month.

Over the past decade, the NHS has strived to promote the use of least RPs in acute mental health services [12,14]. In 2015, the Mental Health Act Code of Practice set predictions for mental health services to reduce restrictive interventions, with this continued shift to reduce the use of RPs clearly outlined in the recent NHS England’s long-term plan [14]. The Care Quality Commission in the United Kingdom independently identified 5 different trusts to be used as an example in programmes implemented to reduce RPs [15]. The report indicated similar approaches within the programs such as improved leadership, restraint reviews, organisational cultural changes, collaborative empowerment, training staff or patients [3,5,6,12,15]. However, research showing the effectiveness of those programs has been scarce, with more detailed studies needed to better understand the issues surrounding RPs.

The development of technology has allowed for new methods of training to be implemented on global scale with knowledge sharing opportunities [16,17]. WEmerging healthcare technologies provide opportunities to simulate virtual environments to better learn and offer an improved understanding of visual stimuli to improve the sensory experiences [17]. One such technology is virtual reality (VR) which can be used as an immersive and engaging teaching tool used for education and training to optimise healthcare outcomes for patients [18]. VR facilitates scenario - based learning to train staff to better manage difficult circumstances such as an encounter with an aggressive or unpredictable patient with complex psychiatric comorbidities. Similar approaches have historically been used within acute clinical areas such as surgery and radiology where healthcare professionals can repeatedly practice and receive feedback on the appropriate response in high pressure situations and reducing the overall risk to the patients and staff [19,20]. VR training has been researched in a variety of healthcare environments since it is a cost-effective and safe method for the practice of heavily-procedure methods, such as the ones used in surgery [17,21]. VR training has been proven to improve the understanding of procedures such as hip replacements [20], laparoscopy [19], screening protocols [17] and elective procedures [21]. By allowing students to practice their training and skills through VR, research has shown promising results in reducing the training curve [20], improved clinical skills [19] and a reduction on human error in surgery allowed by repeated training [22]. It has been claimed that it can change the future of assessment and treatment of numerous disorders in the mental health area [23]. Additionally, VR removes the stress of face-to-face teaching and time constraints in busy hospital environments [21]. In the context of psychiatry, VR tools support healthcare professionals learn from their own approaches in a more controlled setting aiding with developing optimal practice guidelines. Hence, VR has the potential to be an adaptable and realistic tool to improve quality of care offered. In its current form, there are limitations when using VR. VR has the possibility to be expensive as it may require a variety of regular updates limiting its use within publicly funded healthcare organisations such as the NHS and those within low-middle-income countries [16]. Since it is an emerging technology that necessitates ongoing improvements, it is challenging to develop substantial and high-quality evidence [22].

Virti is a digital platform that was created for an immersive learning experience in healthcare simulation and training [24]. The intended purpose of the technology is to facilitate clinical staff training for the practice of clinical and corporate skills. Virti, simulations could generate best practice methods to reduce RPs. User experience data previously gathered on the Virti platform, demonstrates user preference and increased engagement when using the platform over alternatives such as video, audio and online [25]. We designed the tree study to assess the effect of VR training on reducing RPs in acute psychiatric wards.

MATERIALS AND METHODS

Aims and objectives

The primary aim of this feasibility study was to assess the efficacy of using VR scenarios to provide training on reducing RPs within inpatient psychiatric wards. Additionally, the study aimed to assess participant-parameters, in relation to reducing the use of RP based on context-specific confidence, anxiety, compassion and burnout between the intervention and control group. The study also aimed to evaluate the use of the Virti system to deliver reduction in RP training, determining user acceptability, preference, and adherence to the simulations.

In relation to specific measures, the study aimed to assess whether: (1) Participants in the VR group will report higher levels of confidence and compassion than the control group in delivering non-RPs following the training intervention; (2) Participants in the VR group will report lower levels of anxiety, discrimination, and burnout than the control group in delivering non-RPs following the training intervention; and (3) Participants in the VR group will show significant decreases in the use of RPs following the training intervention in comparison to the control group.

Research design

An exploratory, randomised controlled, 2-arm, feasibility trial was designed. The primary site for this study was Southern Health NHS Foundation Trust (SHFT). SHFT has 11 inpatient psychiatric wards, of which 4 are specialist forensic units for both adults and adolescents. Thus, the real-world setting makes this feasibility study more adaptable to mental healthcare settings within the United Kingdom. The following wards took part in the study: (1) Ward 1 - Adolescent Low

Secure Unit; (2) Ward 2 - Adolescent Medium Secure Unit; (3) Ward 3 - CAMHS inpatient; (4) Ward 4 - Adult Acute; (5) Ward 5 - Adult Acute; (6) Ward 6 - Adult Psychiatric Intensive Care Unit (PICU); (7) Ward 7 - Adult Psychiatric Intensive Care Unit; (8) Ward 8 - Adult Low Secure Unit; (9) Ward 9 - Adult Medium Secure Unit; (10) Ward 10 - Old Persons Mental Health (OPMH); and (11) Ward 11 - Old Persons Mental Health. Three wards were randomly allocated the VR training in the intervention arm, with three comparator wards also randomly selected from the pool of remaining inpatient wards. Once allocated, the ward/site managers were approached to discuss the study and implementation of VR training in addition to the mandatory training on reducing RPs for staff working on these wards. Study activities were completed within the participating wards.

Recruitment and selection

Group allocation was clustered at ward level, not at individual level. Study participants were allocated to two parallel study groups. Participation was voluntary. The inclusion criteria for participants were: ≥ 18 years. There was no upper age limit, SHFT employee including: (1) NHS professionals; (2) Clinical staff working in acute inpatient psychiatric wards; (3) Mandatory training on RPs as part of the job role; (4) Willing and able to use VR devices; (5) Ability to give informed consent; and (6) Ability to speak and read English fluently. The exclusion criteria included: (1) SHFT staff not required to complete RPs training; (2) Previous experience on the Virti VR platform; (3) Not willing or unable to give informed consent; (4) Previous experience of cybersickness; and (5) History of epileptic episode (diagnosed or suspected).

Randomisation and blinding

Participants at each randomised ward were invited to take part in the VR study by the SHFT research team. Each eligible staff received an email with a Qualtrics Core XM link to the study. Participants in the intervention arm who completed the consent form were given a login code to access the VR training and use the VR headsets. All participants who consented to the study completed baseline assessments and end of study measures at week four, follow-up measures were conducted at six months.

Intervention

The intervention arm (group A) had completed the Trust's current standard training on reducing RP and then received additional training using the Virti VR system.

Group A: Virti VR training intervention arm

Participants underwent training in patient de-escalation and appropriate use of RPs using VR technology. They were equipped with VR headsets and accessed the training *via* Virti's mobile application. All necessary equipment for the VR training was provided to the participants as part of the study. The VR training comprised four simulations, developed collaboratively by the SHFT Reducing Restrictive Practices training team supported by Virti. Each simulation was designed to last approximately five minutes. Participants received a training guide detailing how to independently complete the course. They were allowed to revisit the training as needed during this period, providing a flexible learning opportunity. Participants were required to complete a minimum of two training sessions for each simulation. The VR simulations incorporated various scenarios, offering perspectives of staff, an observer, and a patient. Four simulations were proposed, and these are described in [Table 1](#).

Group B: Standard training treatment as usual arm

The treatment as usual (TAU) arm only received the Trust's current standard face-to-face and/eLearning training as required by their current job role. The standard training is aimed at clinical staff working within mental health and learning disabilities inpatient services. The training covers the essential aspects of the relevant law and the Trust's standard operating procedures, physical health risks associated with RPs, demonstration of safe and effective application of physical restraints and health monitoring following the intervention.

Outcome measures

Self-efficacy (General self-efficacy scale): The general self-efficacy (GSE) was originally developed in Germany and has been adapted to 28 languages[26]. Numerous studies have demonstrated the GSE to have high reliability, stability, and construct validity[27,28]. The scale includes only one global dimension measured through 10 items. Participants respond to items such as "Thanks to my resourcefulness, I can handle unforeseen situations" using a 4-point Likert scale ranging from 'not at all true' through to 'exactly true'.

Anxiety (Generalised anxiety disorder assessment 7): The generalised anxiety disorder assessment 7 (GAD-7) is a 7-item scale which screens for anxiety and assesses the severity. Each item is scored on a scale of 0, 1, 2, and 3 to the response categories of 'not at all', 'several days', 'more than half the days', and 'nearly every day', respectively, and adding together the scores for the seven questions. Scores of 5, 10, and 15 are taken as the cut-off points for mild, moderate and severe anxiety, respectively. When used as a screening tool, further evaluation is recommended when the score is 10 or greater[29].

Health care professional burnout (Burnout Assessment Tool 12): The Burnout Assessment Tool (BAT)-12 is a concise self-reported questionnaire comprising 12 items across four domains: Exhaustion, mental distance, cognitive impairment, and emotional impairment. Each statement is scored on a scale ranging from 1 (never) to 5 (always)[30].

Table 1 Brief summaries of virtual reality simulations used in the study

Simulation	Scenario	Scene purpose	Scene written description
1	A patient with a weapon approaching a ward staff on the corridor/room - "Chelsea Cosh"/Millwall Brick	To demonstrate de-escalation of situation without the need to use restrictive practices	Staff member encounters patient with a weapon in day room
2	A patient confronting, shouting and screaming at a ward staff	To introduce learner to a spontaneous situation that may occur on a ward	Conversation between colleagues about an escalated patient who is angry that another patient has taken their cigarettes and isn't giving them any back They are distressed that their needs are not being seen to and the ward staff are gesturing that they will be with them soon
3	Experience of being a patient under observation - filmed from a patient's perspective	To give user/learner the experience of being under supervision or observation	Filmed from perspective of patient (1st person). Clinical worker talking to camera explaining that they are being placed under observation
4	Conflict over restricted area. Dealing with conflict over access to a locked fridge/restricted area	For learners to understand/empathise with frustrations around patients trying to do day to day tasks	Set in ward communal area/kitchen. Patient enters and is attempting to make a cup of tea

Discrimination (Everyday Discrimination Scale): The Everyday Discrimination (EDS) is used as a measure of subjective experiences of daily discrimination against the minority population. This measure includes nine elements that evaluate an individual's daily experiences, each followed by a supplementary question probing the perceived reason behind the experienced discrimination[31].

Compassion (Compassionate Engagement and Action Scale): The Compassionate Engagement and Action (CEA) Scales comprises three scales which measure self-compassion. Compassion for others, compassion from others, and compassion for self, each scored separately. For each scale, two subscales: Engagement and actions can be calculated. The questionnaire has two aspects of compassion, the first is the ability to be motivated to engage with things/feelings that are difficult as opposed to trying to avoid or suppress them. The second aspect of compassion is the ability to focus on what is helpful. Participants are asked to rate each statement according to how frequently it occurs on a scale of 1 to 10 (1 being never; 10 being always)[32].

Analysis plan

The baseline questionnaire included sections on demographics and psychosocial factors, drawing from validated psychological scales and ordinal data. Initially, a stepwise approach was employed for data pre-processing, which involved removing repeated samples. Subsequently, we calculated scores for several key measures: Anxiety (GAD-7), GSE, healthcare professional burnout (BAT-12), and the CEA scale for experimenters across all periods. Notably, the BAT-12 score encompasses 12 items across four domains - exhaustion, mental distance, cognitive impairment, and emotional impairment. It also includes additional items measuring the length of working and considerations for seeking other job opportunities, which were analysed on a case-by-case basis. Moreover, the System Usability Scale (SUS) was computed to provide a global view of subjective usability assessments. The evaluation of discrimination was conducted based on quantifiable data.

Initially, hypothesis tests were conducted to assess differences between groups, with an examination of baseline data for psychological indicators. This was followed by comparative analyses of the two subject groups. Subsequent hypothesis testing on the one-month and six-month follow-up data measured the VR intervention's impact. Post-hoc power analysis was then utilised to verify the validity of these hypothesis testing results and to guide sample size power calculation for subsequent trials. Additionally, the SUS score from the one-month data was analysed separately to assess the system's acceptance.

In our hypothesis testing, we treated numerical and ordinal variables distinctly. For numerical variables, we initially applied the Shapiro-Wilk test to assess for normality. This was followed by the Levene test to evaluate homoscedasticity between the two data groups, determining the appropriate version of the two-sample *t*-test to use. In cases where normality was not satisfied, we employed the Wilcoxon rank sum test to compare the medians of the two groups. For ordinal variables, the Wilcoxon rank sum test was the method of choice.

Hypothesis tests and power analysis

We had two groups of data $\{y1j\} = 1n1$ in the VR group, and $\{y2j\} = 1n2$, in the standard training group, where $y1j$ and $y2j$ were values of our indicator of interest, and $n1$ and $n2$ were sample sizes.

In the two-sample *t*-test assuming homoscedasticity, the null hypothesis was $H0:M1 = \mu2$, where μ_i represented the mean of the indicator (e.g., GAD7 score, GSE score, etc.) for group i , and the alternative hypothesis was $H1:M1 \neq \mu2$, where $y1$ and $y2$ represented the means of the data for both groups. The sample variances $S12$ and $S22$ were calculated as $S12 = 1n1 - 1j = 1n1(y1j - y1)^2$ and $S22 = 1n2 - 1j = 1n2(y2j - y2)^2$, respectively. The test statistic below was derived from these values:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Under the null hypothesis H_0 , the test statistic followed a Student's- t distribution with $n_1 + n_2 - 2$ degrees of freedom, denoted as T . From the given data, we calculated the observed t -value, represented as t_{obs} . The P value was then determined by $P = 2\Pr(T \geq |t_{obs}|)$. If $P < 0.05$, we rejected H_0 , indicating sufficient evidence to conclude a significant difference between the means of the two groups at a 0.05 confidence level. If $P \geq 0.05$, we would not have strong evidence to support this conclusion.

For the Wilcoxon rank sum test, the null hypothesis $H_0: M_1 = m^2$ where m_i denotes the median of group i . The alternative hypothesis was $H_1: M_1 \neq m^2$, with R_i the rank of the i th observation in the combined sample, where $1 \leq i \leq n_1 + n_2$. We defined I_i as 1 if the i th observation came from group 1, and 0 if it came from group 2. The test statistic was then defined as:

$$W = \sum_{i=1}^{n_1+n_2} R_i I_i$$

Given that both n_1 and n_2 were greater than 10, the distribution of the Wilcoxon rank sum test statistic W was approximated using the standard normal distribution.

Under the null hypothesis H_0 , the test statistic W was normalised as follows:

$$Z = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}} \rightarrow N(0, 1)$$

Using the data, we calculated the observed value of test statistic, denoted as z_{obs} . The P value was then determined by $P = 2\Pr(Z \geq |z_{obs}|)$, where Z followed as standard normal distribution, $N(0,1)$. If $P < 0.05$, we rejected the null hypothesis H_0 , implying sufficient evidence to conclude a significant difference between the medians of the two groups at the 0.05 significance level. If $P \geq 0.05$, there was insufficient evidence to reject H_0 . In addition to calculating P values, a *post-hoc* power analysis was conducted to evaluate the reliability of our conclusions and to guide future studies. This will hopefully guide the power calculations for subsequent studies.

RESULTS

Demographic information for participants

Figure 1 presents the flow of participants into the trial. Demographic data was gathered for all patients and is represented in Table 2. Total of 56 participants took part in the study, allocated to either VR group ($n = 34$) or TAU group ($n = 22$).

Comparison between two groups for baseline data

As illustrated in Table 3, the P values for most attributes exceed 0.05, with the test powers being large for the majority of these variables. Exceptions include the BAT12 score, compassionate engagement to others' score, compassionate to others' total score, compassionate engagement from others' score, and compassionate from others' total score, all analysed using two-sample t -tests. This may be attributed to the inherent characteristics of the t -test. To increase statistical power, it will be necessary to enlarge the sample size in future iterations of the experiment. Overall, the current analysis suggests that there is no significant difference between the groups based on the P values. However, for BAT 12 item "Thinking back from this moment, over the past 24 hours, how many hours have you worked in your job?", EDS item 7, and EDS item 9, the P values are below 0.05, suggesting potential significant differences between the two groups in these measurements. Nonetheless, given that these differences pertain to only a few aspects in BAT12, it can be inferred that there are essentially no significant differences between the two groups overall. This observation is consistent with the outcomes commonly expected in randomised controlled trials.

We evaluated various psychosocial indicators and rates of RP. Hypothesis testing was not applied to the 6-month data due to the small sample sizes, with only 8 samples in the VR group and 5 in the control group. This limited data set could lead to high bias, rendering the testing results unreliable. Therefore, the 6-month data were solely analysed through frequency analysis and are presented in the [Supplementary material](#).

User acceptability, preference, and adherence to the program

For the evaluation, we utilised the SUS questionnaire. Participants in the VR group completed this questionnaire after one month. Initially, the responses to all questions on the SUS questionnaire are depicted in Figure 2. These responses are ordered as per Likert scale from 0 being strongly disagree to 5 being strongly agree. To assess the use of Virti VR technology in the reduction in RP training program and to determine user acceptability, preference, and adherence, we computed the SUS scores. Pie charts were also created to visualise the completion rate of the training program (Figure 3). The histogram displaying the SUS scores is illustrated in Figure 4. The majority of respondents reported SUS scores above 70, with an average score of 71.79. According to Bangor *et al*'s findings, this positions the VR platform as more usable than approximately 62% of other products[33].

GAD7 comparative analysis

The GAD7 score was calculated to evaluate the confidence levels between the two groups, and hypothesis tests were applied to the data. Power analysis was conducted to validate the P value findings. We compared the data after one month for both the VR group and the control group, as well as the baseline and one-month data for each group. The

Table 2 Demographic characteristics of the sample

Characteristic	VR (n = 34)	TAU (n = 22)
Age (year), mean (SD)	40.8 (13.2)	40.7 (11.9)
Sex, n (%)		
Male	9 (26.5)	10 (45.5)
Female	25 (73.5)	12 (54.5)
Ethnicity, n (%)		
White	23 (67.6)	11 (50)
Other white background	2 (5.9)	1 (4.5)
White and Asian	1 (2.9)	1 (4.5)
Indian	1 (2.9)	0 (0)
Pakistani	0 (0)	1 (4.5)
Other Asian background	1 (2.9)	0 (0)
African	4 (11.8)	8 (36.4)
Other Black background	1 (2.9)	0 (0)
Other	1 (2.9)	0 (0)
Education level, n (%)		
No formal qualifications	1 (2.9)	0 (0)
GCSE or equivalent	9 (26.5)	4 (18.2)
A-level or equivalent	7 (20.5)	4 (18.2)
Undergraduate	15 (44.1)	9 (40.9)
Postgraduate or above	2 (5.9)	5 (22.7)
Disorder, n (%)		
Anxiety	1 (2.9)	2 (9.1)
Depression	8 (23.5)	4 (18.2)
Post-traumatic stress disorder	1 (2.9)	0 (0)
None	21 (61.8)	13 (59.1)
Other	3 (8.8)	3 (13.6)
Role, n (%)		
Mental health nurse	9 (26.5)	9 (40.9)
Healthcare support worker	22 (64.7)	12 (54.5)
Occupational therapist	0 (0)	1 (4.5)
Other	3 (8.8)	0 (0)
Work time in NHS, n (%)		
> 8 months	0 (0)	1 (4.5)
8 months to 1 years	3 (8.8)	1 (4.5)
1-5 years	19 (55.9)	6 (27.3)
6-10 years	2 (5.9)	2 (9.1)
< 10 years	10 (29.4)	12 (54.5)
Experience in inpatient psychiatric wards, n (%)		
> 1 year	3 (8.8)	2 (9.1)
1-2 years	5 (14.7)	2 (9.1)
3-4 years	7 (20.6)	1 (4.5)

4-5 years	7 (20.6)	3 (13.6)
≤ 5 or more year	12 (35.3)	14 (63.6)
Experience in restrictive training, <i>n</i> (%)		<i>n</i> = 21
None	2 (5.9)	1 (4.5)
> 1 year	3 (8.8)	6 (27.3)
1-2 years	6 (17.6)	1 (4.5)
3-4 years	4 (11.8)	0 (0)
4-5 years	6 (17.6)	3 (13.6)
≤ 5 or more years	13 (38.2)	10 (45.5)

VR: Virtual reality; TAU: Treatment as usual; GCSE: General Certificate of Secondary Education; NHS: National Health Service.

Table 3 Baseline comparison

Characteristic	VR (<i>n</i> = 34)	TAU (<i>n</i> = 21)	<i>P</i> value	Effective size	Power
Total GAD7 score, mean (SD)	2.88 (4.13)	4.32 (5.71)	0.663	0.0597	0.9998
Total GSE score, mean (SD)	32.71 (4.36)	31.14 (3.73)	0.295	0.142	0.9939
EDS measure					
Item 1	4.29 (1.27)	3.71 (1.52)	0.166	0.194	0.9685
Item 2	4.50 (1.13)	3.90 (1.37)	0.93	0.205	0.9575
Item 3	5.53 (0.71)	5.43 (1.03)	0.986	0.0028	1
Item 4	4.79 (1.07)	5.14 (1.11)	0.209	0.179	0.9796
Item 5	5.35 (0.88)	4.71 (1.74)	0.363	0.136	0.9951
Item 6	5.62 (0.70)	5.71 (0.56)	0.722	0.0616	0.9998
Item 7	4.88 (0.98)	4.29 (0.90)	0.035 ^a	0.296	0.7211
Item 8	5.06 (1.37)	4.29 (1.79)	0.096	0.239	0.9028
Item 9	5.09 (1.22)	3.71 (1.79)	0.005 ^b	0.395	0.258
Total BAT12 score	27.24 (6.99)	25.67 (7.00)	0.423	0.2211	0.1225
Compassionate					
To others	78.03 (11.65)	82.86 (8.36)	0.109	0.4458	0.3511
From others	68.56 (18.65)	67.62 (24.65)	0.874	0.0435	0.0527
Compassion engagement					
To others	42.97 (8.65)	46.67 (6.97)	0.101	0.4564	0.3651
From others	37.94 (11.92)	39.14 (14.79)	0.742	0.0907	0.0619
Compassionate action					
To others	35.06 (5.07)	36.19 (4.34)	0.377	0.121	0.9972
From others	30.62 (7.82)	28.48 (10.30)	0.55	0.0808	0.9995

^a*P* < 0.05.

^b*P* < 0.01.

VR: Virtual reality; TAU: Treatment as usual; GAD7: Generalised anxiety disorder assessment 7; GSE: General self-efficacy; EDS: Everyday Discrimination Scale; BAT12: Burnout Assessment Tool 12.

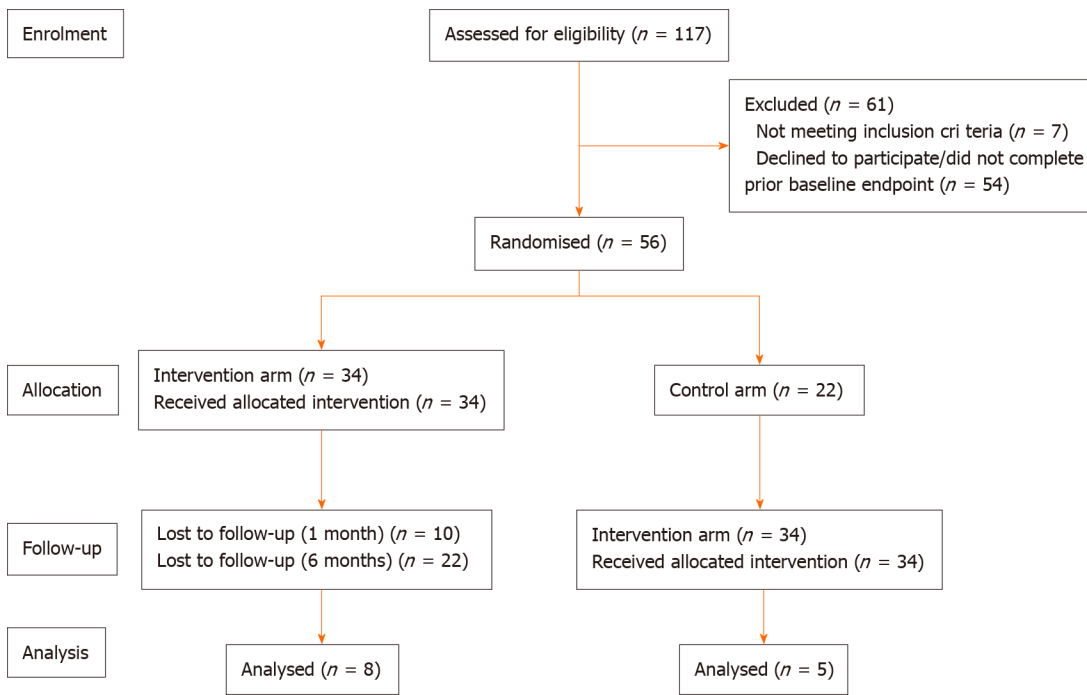


Figure 1 Study consort flow diagram.

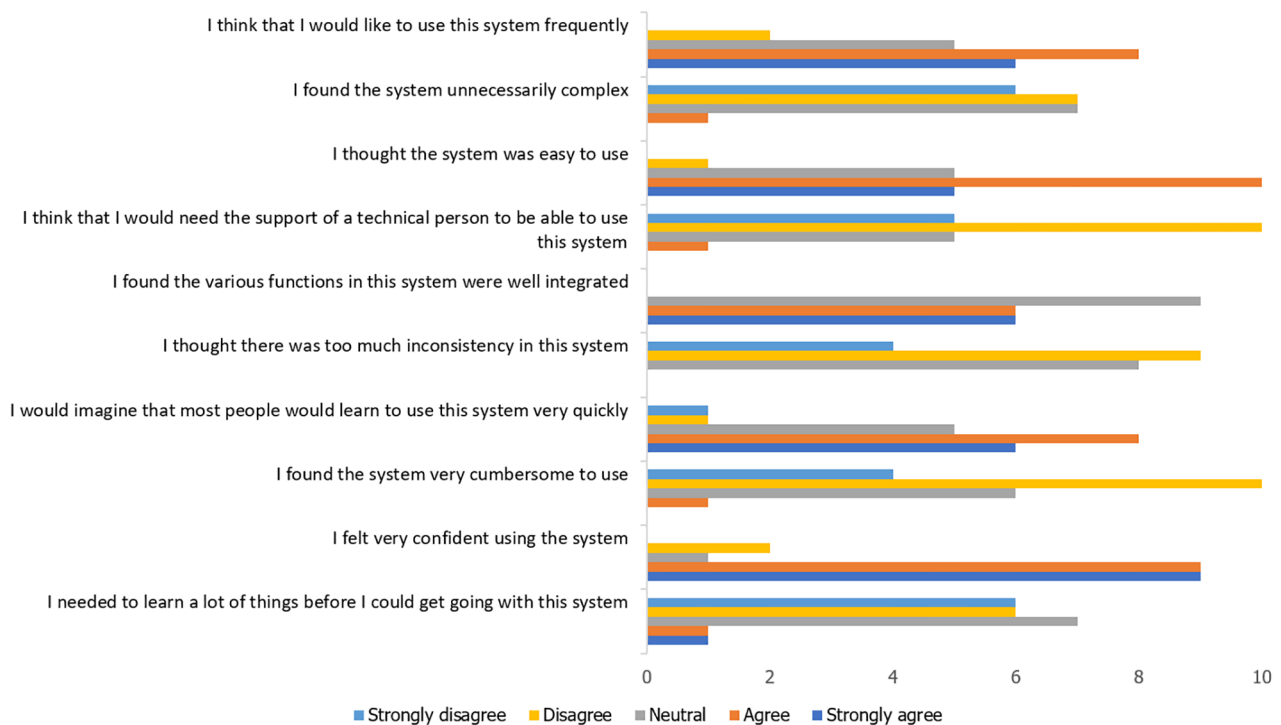


Figure 2 Frequency analysis of the System Usability Scale questionnaire.

GAD7 scores across different time points in the study found no statistically significant differences when comparing the VR group and the TAU group. Specifically, the VR group at one-month follow-up ($n = 23$) had a mean (SD) = 3.39 (3.60), compared to TAU ($n = 17$), mean (SD) = 5.24 (5.56); the P value of 0.702 indicates a lack of significant difference at this time point. When assessing the VR group across baseline and one-month data [$n = 23$ with mean (SD) = 3.39 (3.60) *vs* $n = 34$ with mean (SD) = 2.88 (4.13); $P = 0.341$], suggests that changes over time within the VR group were also not significant. A similar pattern was observed in the TAU group, with no significant change in GAD7 scores between baseline and one-month follow-up [$n = 17$ with mean (SD) = 5.24 (5.56) *vs* $n = 22$ with mean (SD) = 4.32 (5.17); $P = 0.64$]. The power analysis yielded high values (greater than 0.95) for all comparisons, suggesting that the sample size was adequate to detect differences if present. The effect sizes were small across all comparisons, implying minimal clinical differences between

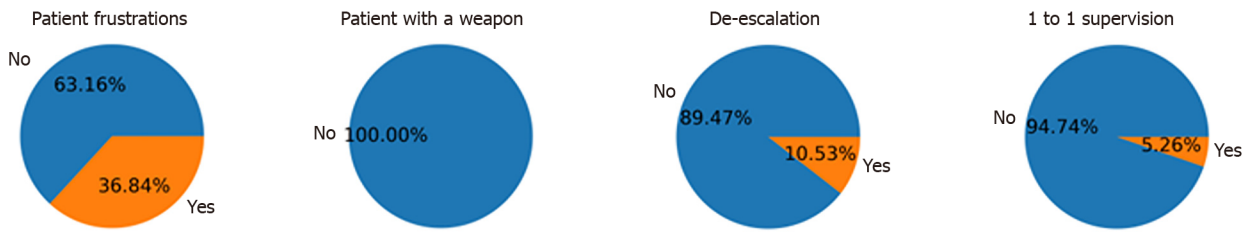


Figure 3 The distribution of the experimenter’s completion.

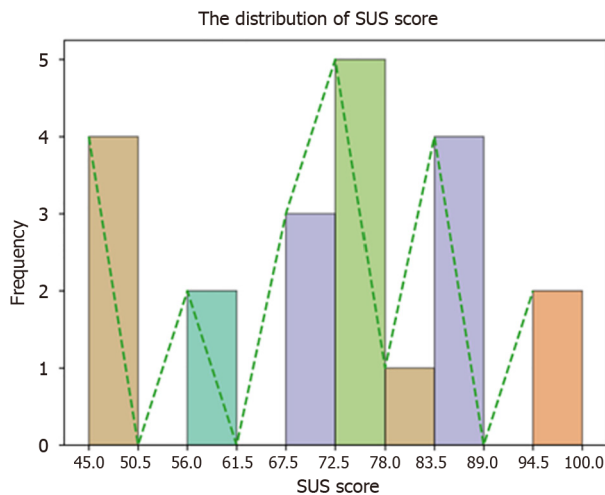


Figure 4 The distribution of the System Usability Scale score. SUS: System Usability Scale.

groups and across time points (Table 4).

GSE comparison of levels

Table 4 shows that the comparison of GSE scores between the VR group and the TAU group indicated no statistically significant differences at any measured time point. For the one-month data, the VR ($n = 23$) exhibited mean (SD) = 31.78 (4.31), while the TAU ($n = 17$) had a mean (SD) = 32.82 (4.75); $P = 0.367$. This suggests that the VR intervention did not significantly affect self-reported anxiety levels compared to the TAU group. The effect size was small (0.143), and the statistical power of the test was high (0.9717), indicating a sufficient sample size for detecting a real difference if one existed. Further comparison within the VR group between baseline and one-month follow-up data mean (SD) = 31.78 (4.31) vs mean (SD) = 32.71 (4.36); $P = 0.531$, reinforcing the finding of no significant change over time. The effect size for this comparison was very small (0.0833), and the power was near perfect (0.9996).

Similarly, comparing the TAU group’s baseline and one-month follow-up data ($n = 17$; mean (SD) = 32.82 (4.75) vs $n = 21$; mean (SD) = 31.14 (4.73); $P = 0.147$. Here, the effect size was modest (0.237), but the power was somewhat lower (0.7974), which could suggest a reduced ability to detect differences. Overall, the VR intervention did not significantly impact the self-efficacy levels as measured by the GSE, compared to the TAU group. Additionally, within-group changes over time were not statistically significant for both groups. These conclusions are drawn from the context of sufficient statistical power, indicating the sample size albeit small was adequate.

Comparison of levels of discrimination

The EDS was employed to measure levels of discrimination. Hypothesis tests were conducted for each question, followed by a power analysis. A comparative analysis of the day-30 data was then carried out between the VR group and the TAU group (Table 5). In examining the levels of perceived discrimination using the EDS, our analysis revealed that, generally, there were no statistically significant differences between the VR and TAU groups across most elements measured. Notably, at the one-month follow-up, the VR group ($n = 23$) and the TAU group ($n = 17$) showed no significant differences in scores for questions Q1 through to Q7, with P values ranging from 0.286 to 0.753, suggesting similar levels of perceived daily discrimination in both groups. The effect sizes for these questions were generally small, with most below 0.2, indicating a minor practical significance of the findings.

However, for item Q8 (You are called names or insulted), the VR group reported mean (SD) = 5.17 (1.11), while in the TAU group, mean (SD) = 4.06 (1.60); $P = 0.023$. This suggests that there is a statistically significant difference between the VR and TAU groups for this item, with the VR group reporting higher levels of perceived discrimination. The effect size is 0.375, a moderate effect size, and the power of the test is 0.2722, which is relatively low. This low power means there is a higher chance that the difference could be due to random variability rather than a true effect of the intervention.

Table 4 Levels of confidence and anxiety

Outcome measures	Sample groups	Baseline	One-month	P value	Effectiveness size	Power
GAD	VR (<i>n</i> = 23), mean (SD)	2.88 (4.13)	3.39 (3.60)	0.341	0.128	0.9577
	TAU (<i>n</i> = 17), mean (SD)	4.32 (51.7)	5.24 (5.56)	0.64	0.0768	0.9951
	VR vs TAU ¹	-	-	0.702	0.0613	0.9974
GSE	VR (<i>n</i> = 23), mean (SD)	32.17 (4.13)	31.78 (4.31)	0.531	0.0833	0.9996
	TAU (<i>n</i> = 17), mean (SD)	31.14 (3.73)	32.83 (4.75)	0.147	0.237	0.7974
	VR vs TAU ¹	-	-	0.367	0.143	0.9717

¹Virtual reality vs treatment as usual comparison at one-month.

VR: Virtual reality; TAU: Treatment as usual; GAD: Generalised anxiety disorder assessment; GSE: General self-efficacy.

The most substantial difference was observed in item Q9 (You are threatened or harassed) at the one-month follow-up for the VR group, with a mean score mean (SD) = 5.13 (1.14) and the TAU group mean (SD) = 4.18 (1.51); $P = 0.040$, a statistically significant difference between the two groups, with the VR group again reporting higher levels of perceived discrimination. The effect size is 0.339, indicating a moderate effect size. However, the power for this test is 0.4144, which is also considered low, suggesting caution in interpreting this result as definitive evidence of a real difference between groups due to the potential risk of type II error.

Within the VR group's baseline and one-month follow-up comparisons, as well as the TAU group's comparisons, no significant differences were found for most items, with P values well above the 0.05 threshold. The effect sizes remained small, and the power of these tests was generally high, exceeding 0.95, indicating a robust ability to detect significant effects if they were present.

In summary, while the study found minimal evidence of significant differences in perceived discrimination between the VR and TAU groups, two items (Q8 and Q9) in the one-month follow-up data suggest areas where the VR intervention might influence reported discrimination experiences. These preliminary findings highlight the need for further research with larger sample sizes and more refined power to fully understand the potential impact of VR interventions on perceptions of discrimination.

Burnout

In an analysis of burnout levels using the BAT-12 scale (Table 6), the study found no statistically significant differences in overall burnout scores or individual item scores between the VR group and the TAU group. Specifically, the VR group at one-month follow-up ($n = 23$) reported a mean (SD) overall BAT-12 score of 29.13 (7.25), while the TAU group ($n = 16$) reported a mean (SD) score of 26.94 (7.64); $P = 0.369$ indicating no significant difference between the groups.

When examining individual items of the BAT-12, P values ranged from 0.253 to 0.797 for the VR group and from 0.075 to 0.794 for the TAU group, all above the conventional threshold of 0.05 for statistical significance. This indicates a consistent lack of significant difference in perceptions of exhaustion, mental distance, cognitive impairment, and emotional impairment as measured by the BAT-12 between the VR and TAU groups. Additionally, within the VR group ($n = 23$ at one month, compared to $n = 34$ at baseline) and the TAU group ($n = 16$ at one month, compared to $n = 21$, baseline), the changes in burnout scores over time did not reveal any statistically significant differences, with P values well above 0.05.

The effect sizes across all comparisons were generally small, with few exceptions where moderate effect sizes were observed (e.g., Q24 in the TAU group comparison). Nevertheless, these did not translate into statistically significant differences. The statistical power for most comparisons was high, suggesting that the sample sizes were adequate to detect significant differences if they were present. Overall, the findings suggest that the intervention with VR did not have a statistically significant impact on burnout levels compared to the TAU group. The absence of significant differences in burnout levels, as well as the individual elements of the BAT-12, may indicate that both interventions had a comparable influence on the participants' burnout symptoms.

Compassion

To assess compassion levels, we calculated the compassionate engagement, action, and total scales from two parts of the questionnaire. These two parts respectively measured the level of compassion towards others and the level of compassion received from others. From the data, it is apparent that there were no statistically significant differences between the VR and TAU groups in terms of compassionate engagement to others, compassionate action to others, and their total scores. The P values for these comparisons are above the conventional threshold of 0.05, indicating a lack of significant difference. This is supported by the effect sizes, which are mostly small, suggesting any observed differences are minimal in practical terms (Table 7).

Notably, there was one significant result within the VR group for compassionate engagement from others ($P = 0.039$), indicating a significant difference between the two time points for this component. However, the effect size is moderate (0.6839), and the power for this test is relatively low (0.5344), suggesting that while a significant difference was observed, the reliability of this result may be questionable due to potential type II error risks.

Table 5 Levels of discrimination - Everyday Discrimination Scale

Outcome measures	Sample groups	Baseline	One-month	P value	Effectiveness size	Power
Item 1	VR (<i>n</i> = 23), mean (SD)	4.29 (1.27)	4.13 (1.29)	0.62	0.0677	0.9998
	TAU (<i>n</i> = 17), mean (SD)	3.71 (1.52)	4.12 (1.54)	0.454	0.125	0.9784
	VR vs TAU ¹	-	-	0.924	0.0155	0.9995
Item 2	VR (<i>n</i> = 23), mean (SD)	32.17 (4.13)	31.78 (4.31)	0.531	0.0833	0.9996
	TAU (<i>n</i> = 17), mean (SD)	31.14 (3.73)	32.83 (4.75)	0.147	0.237	0.7974
	VR vs TAU ¹	-	-	0.367	0.143	0.9717
Item 3	VR (<i>n</i> = 23), mean (SD)	5.53 (0.71)	5.08 (10.4)	0.152	0.213	0.9575
	TAU (<i>n</i> = 17), mean (SD)	5.43 (10.3)	5.47 (0.80)	0.918	0.0198	0.9992
	VR vs TAU ¹	-	-	0.286	0.187	0.9236
Item 4	VR (<i>n</i> = 23), mean (SD)	4.79 (1.07)	4.78 (1.24)	0.909	0.0157	1
	TAU (<i>n</i> = 17), mean (SD)	5.14 (1.03)	5.06 (1.03)	0.725	0.0619	0.9964
	VR vs TAU ¹	-	-	0.547	0.1	0.9912
Item 5	VR (<i>n</i> = 23), mean (SD)	5.35 (0.88)	5.09 (1.08)	0.375	0.129	0.9974
	TAU (<i>n</i> = 17), mean (SD)	4.71 (1.74)	5.06 (1.43)	0.628	0.0864	0.9925
	VR vs TAU ¹	-	-	0.692	0.0678	0.9967
Item 6	VR (<i>n</i> = 23), mean (SD)	5.62 (0.70)	5.30 (0.97)	0.216	0.192	0.9768
	TAU (<i>n</i> = 17), mean (SD)	5.71 (0.56)	5.59 (0.71)	0.714	0.0773	0.9943
	VR vs TAU ¹	-	-	0.359	0.166	0.9513
Item 7	VR (<i>n</i> = 23), mean (SD)	4.88 (0.98)	4.43 (1.34)	0.251	0.157	0.9927
	TAU (<i>n</i> = 17), mean (SD)	4.29 (0.90)	4.35 (0.86)	0.86	0.0306	0.9988
	VR vs TAU ¹	-	-	0.753	0.0517	0.9981
Item 8	VR (<i>n</i> = 23), mean (SD)	5.06 (1.37)	5.17 (1.11)	0.981	0.0036	1
	TAU (<i>n</i> = 17), mean (SD)	4.29 (1.79)	4.06 (1.60)	0.567	0.095	0.9904
	VR vs TAU ¹	-	-	0.023	0.375	0.2722
Item 9	VR (<i>n</i> = 23), mean (SD)	5.09 (1.22)	5.13 (1.14)	0.987	0.0024	1
	TAU (<i>n</i> = 17), mean (SD)	3.71 (1.79)	4.18 (1.51)	0.445	0.126	0.9778
	VR vs TAU ¹	-	-	0.04	0.339	0.4144

¹Virtual reality vs treatment as usual comparison at one-month.

VR: Virtual reality; TAU: Treatment as usual.

When examining the within-group comparisons for the VR and TAU groups separately, again no significant differences were found for most components, with all *P* values well above 0.05, reinforcing the lack of significant change over time. The Wilcoxon *P* values and power for these comparisons, where reported, generally support these findings, with most power values indicating a high probability of detecting a true effect if one existed.

In summary, the results indicates that the VR intervention did not significantly influence levels of compassion towards or from others, compared to TAU. However, the significant change in compassionate engagement from others within the VR group suggests a potential area of impact, which may warrant further investigation with a larger sample size and more robust power to confirm these preliminary findings.

RP use

Figure 5 illustrates the comparative rates of RPs between the VR intervention group and the TAU group over the study period. The VR intervention group's RP rates at baseline were higher in comparison to TAU group. These reduced within the 30-day period, the VR group experienced more fluctuation in RP rates likely due to other ward dynamics including staff turnover, not influenced by VR intervention. Notably, the RP rates in the VR group ended on a downward trend. Conversely, the TAU group exhibited consistent RP rates with a steady decline over the same period, possibly indicating the effectiveness of existing standard practices or a different aspect of the ward's dynamics not influenced by the VR intervention. The data suggest a potential for VR interventions to contribute to the reduction of RPs in a clinical setting,

Table 6 Levels of burnout - Burnout Assessment Tool 12 Scale

Outcome measures	Sample groups	Baseline	One-month	P value	Effectiveness size	Power
Total BAT12 score	VR (<i>n</i> = 23), mean (SD)	27.24 (6.99)	29.13 (7.25)	0.625	0.2634	0.1601
	TAU (<i>n</i> = 17), mean (SD)	25.67 (7.00)	26.94 (7.64)	0.602	0.1708	0.0792
	VR <i>vs</i> TAU ¹	-	-	0.369	0.2899	0.1397
Item 1	VR (<i>n</i> = 23), mean (SD)	3.41 (1.79)	2.78 (1.54)	0.157	0.195	0.9746
	TAU (<i>n</i> = 17), mean (SD)	2.19 (1.54)	3.19 (1.76)	0.075	0.307	0.5218
	VR <i>vs</i> TAU ¹	-	-	0.408	0.136	0.9721
Item 2	VR (<i>n</i> = 23), mean (SD)	3.91 (1.11)	3.52 (1.31)	0.302	0.145	0.9953
	TAU (<i>n</i> = 17), mean (SD)	3.67 (1.32)	3.81 (1.28)	0.794	0.0453	0.9973
	VR <i>vs</i> TAU ¹	-	-	0.432	0.132	0.9748
Item 3	VR (<i>n</i> = 23), mean (SD)	1.18 (0.39)	1.22 (0.42)	0.795	0.0504	0.9999
	TAU (<i>n</i> = 17), mean (SD)	1.10 (0.30)	1.00 (0)	0.624	0.206	0.8663
	VR <i>vs</i> TAU ¹	-	-	0.253	0.316	0.4992
Item 4	VR (<i>n</i> = 23), mean (SD)	1.85 (0.93)	1.74 (0.62)	0.916	0.0151	1
	TAU (<i>n</i> = 17), mean (SD)	2.29 (0.90)	2.00 (0.73)	0.391	0.151	0.9538
	VR <i>vs</i> TAU ¹	-	-	0.304	0.183	0.9212
Item 5	VR (<i>n</i> = 23), mean (SD)	2.03 (1.06)	2.04 (0.88)	0.801	0.0351	1
	TAU (<i>n</i> = 17), mean (SD)	1.81 (1.17)	1.94 (1.18)	0.646	0.0824	0.9918
	VR <i>vs</i> TAU ¹	-	-	0.458	0.126	0.9783
Item 6	VR (<i>n</i> = 23), mean (SD)	2.06 (0.95)	2.09 (1.12)	0.961	0.0069	1
	TAU (<i>n</i> = 17), mean (SD)	1.81 (1.08)	2.06 (1.29)	0.581	0.098	0.9873
	VR <i>vs</i> TAU ¹	-	-	0.797	0.0433	0.9981

¹Virtual reality *vs* treatment as usual comparison at one-month.

VR: Virtual reality; TAU: Treatment as usual; BAT: Burnout Assessment Tool.

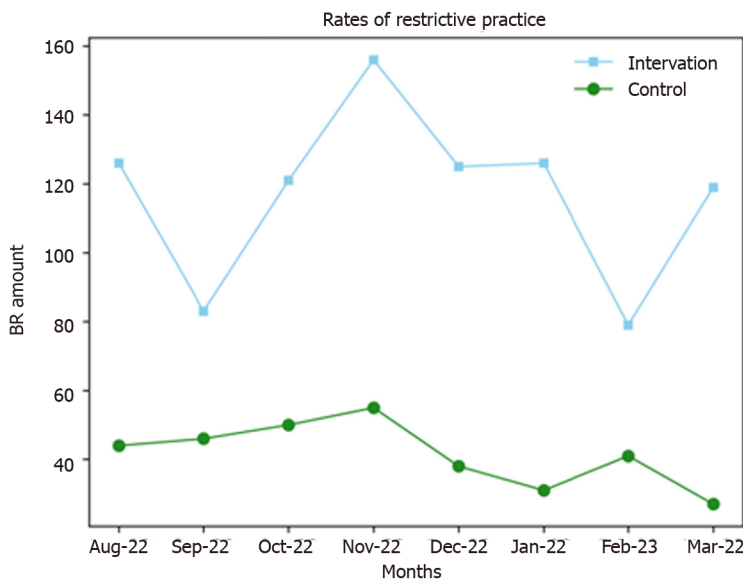


Figure 5 Rates of restrictive practice. RP: Restrictive practice.

Table 7 Levels of compassion - the Compassionate Engagement and Action Scales

Outcome measures	Sample groups	Baseline	One-month	P value	Effectiveness size	Power
Total compassionate score						
To others	VR (<i>n</i> = 23), mean (SD)	78.03 (11.65)	76.26 (9.64)	0.548	0.1611	0.0902
	TAU (<i>n</i> = 17), mean (SD)	82.86 (8.36)	80.69 (6.81)	0.424	0.2624	0.1201
	VR vs TAU ¹	-	-	0.122	0.5041	0.326
From others	VR (<i>n</i> = 23), mean (SD)	68.56 (18.65)	62.36 (15.30)	0.205	0.3464	0.2375
	TAU (<i>n</i> = 17), mean (SD)	67.62 (24.65)	73.13 (22.81)	0.492	0.2256	0.1014
	VR vs TAU ¹	-	-	0.051	0.5605	0.3824
Compassionate engagement						
To others	VR (<i>n</i> = 23), mean (SD)	42.97 (8.65)	42.21 (6.89)	0.728	0.093	0.0632
	TAU (<i>n</i> = 17), mean (SD)	46.67 (6.07)	44.13 (4.21)	0.177	0.4472	0.2586
	VR vs TAU ¹	-	-	0.331	0.3141	0.1557
From others	VR (<i>n</i> = 23), mean (SD)	37.94 (11.92)	34.65 (9.32)	0.270	0.2965	0.1904
	TAU (<i>n</i> = 17), mean (SD)	39.14 (14.79)	42.38 (13.21)	0.495	0.2237	0.1005
	VR vs TAU ¹	-	-	0.039 ^a	0.6839	0.5344
To others	VR (<i>n</i> = 23), mean (SD)	35.06 (5.07)	34.04 (5.05)	0.375	0.1180	0.9984
	TAU (<i>n</i> = 17), mean (SD)	36.19 (4.34)	36.56 (4.94)	0.679	0.0694	0.9944
	VR vs TAU ¹	-	-	0.058	0.307	0.5373
From others	VR (<i>n</i> = 23), mean (SD)	30.62 (7.82)	27.64 (7.31)	0.138	0.199	0.9629
	TAU (<i>n</i> = 17), mean (SD)	28.48 (10.30)	30.75 (10.36)	0.540	0.101	0.9863
	VR vs TAU ¹	-	-	0.156	0.232	0.8066

^a*P* < 0.05.¹Virtual reality vs treatment as usual comparison at one-month.

VR: Virtual reality; TAU: Treatment as usual.

but also highlight the need for a more in-depth understanding of how such interventions can be optimised for consistency and efficacy in real-world applications.

DISCUSSION

VR, as an immersive and engaging intervention, has been recognised for its potential to facilitate learning without compromising patient or staff safety[19-22]. In clinical settings, particularly among orthopaedic teams, VR training has been favourably received. It's noted that VR may become a standardised practice due to its ability to mimic real-time procedures, coupled with the convenience and accessibility it offers compared to traditional training methods[34]. Our feasibility study aimed to evaluate the usability, impact of VR intervention on self-efficacy, anxiety, burnout, and compassion within psychiatric inpatient clinical settings. The findings indicated higher levels of perceived discrimination reported by the VR group when compared to the TAU, the low power for these tests means these results should be interpreted with caution. More robust studies with larger sample sizes might be necessary to confirm these findings.

There were no significant differences between the VR and TAU groups in terms of GSE scores, suggesting that the VR intervention did not influence self-efficacy in this sample in a clinically meaningful way. Anxiety levels, as measured by the GAD7 scale, also did not differ significantly between groups. This may be due to the brief intervention period or the nature of the VR simulation content, which may not have been sufficiently tailored to address specific anxiety-inducing stimuli. Moreover, given the high prevalence and multifaceted nature of anxiety in ward milieu, the absence of change could reflect the resilience of existing anxiety levels to short-term interventions. Similarly, burnout levels assessed by the BAT12 scale were not significantly different between the VR and TAU groups. This finding could also suggest that the VR intervention, while innovative, may not address all the contributing factors to burnout, such as workplace culture, job demands, and individual coping strategies.

The compassion scores, comprising both self-compassion and compassion for and from others, provided a nuanced view. While no significant differences were found between the VR and TAU groups, a significant difference was observed

within the VR group for compassionate engagement from others over time. This isolated finding suggests a potential for VR to influence perceptions of receiving compassion, albeit this result should be interpreted with caution due to the low power of the test. The implication that VR could be a medium through which individuals perceive enhanced compassion from others is an intriguing possibility that warrants further exploration.

Another noteworthy significant finding was the reduction in RPs in the VR group during the intervention study period, laying the groundwork for a more extensive study with a larger sample size. The observed variability in the VR group could be due to multiple factors, including the novelty of the VR training, varying degrees of staff engagement, staff turnover, or external influences not captured in the study. While the eventual decrease in RP rates for the VR group is encouraging, the observed fluctuations warrant a closer examination of the implementation process and the intervention's integration into routine practice.

It appears that, in comparison to conventional training, VR training is more acceptable to staff. This preference may stem from the immersive and engaging nature of VR, making the training more relatable and practical. Overall, VR training seems to yield better outcomes and experiences for staff, even when compared to standard training scenarios. If VR training proves highly effective and applicable in reducing RPs, it could be integrated into regular training schedules, revolutionising traditional training approaches. Furthermore, accessibility of the training at a time convenient to the staff, mean time taken travelling to training centres for face-to-face training are reduced and there is ability for regular refreshers and keeping up to date.

Limitations

The absence of widespread significant findings might be attributed to several factors, including the relatively short duration of the VR intervention, the small sample size, single centre, or the ceiling effects within the chosen psychosocial measures. High power in most tests indicates that the sample size was sufficient to detect differences if they were present, suggesting that any real differences made by the VR intervention were likely small and of limited practical significance. To validate these findings further, a clinical trial using multiple sites would be required. However, the results indicate the *Virti* system could be a valuable tool for hospital settings. Previous research that implemented VR training as part of healthcare education found VR as an efficient way to improve knowledge but also increased satisfaction with the training [22]. Research has shown that VR training leads to modest improvements in knowledge acquisition and significant enhancements in skill development when compared to standard training methods [23]. However, it's important to note that these improvements in knowledge and skills gained through VR training may not directly translate into enhanced clinical competencies [23]. Another limitation is that many studies on VR program implementation have been context-specific and not integrated into standard practices, making them niche to particular environments and situations. Despite these challenges, the usability of the *Virti* VR training system has been evaluated as better than 62% of similar products, which underscores its relevance and potential applicability in training for reducing RPs [35]. The recruitment numbers for our study were low, and we observed a low rate of training completion. This could contribute to the lack of significant findings observed. As larger sample sizes typically provide greater statistical strength and are more effective in determining causal relationships [36], the outcomes of this study might be considered less significant due to this limitation. Such challenges are often attributed to the high-pressure and understaffing issues prevalent in ward environments. Consequently, staff members are frequently overwhelmed with their routine duties, leaving them with little capacity to participate in research studies.

Increasing the interactivity of VR simulators and expanding the range of use-case scenarios would be highly beneficial, particularly if they allow staff to be more actively involved in decision-making processes during the training. Such an approach would foster a more proactive and engaging learning experience. Many VR simulations currently offer users the ability to control and manipulate scenarios, enhancing their educational impact. However, while our study utilised a 360-degree immersive environment, it lacked opportunities for participants to actively engage with the programme. Incorporating interactive elements would significantly enhance the training effectiveness by allowing users to practice decision-making in a simulated, yet realistic, setting.

Implications for future research

VR training is increasingly being adopted across various healthcare disciplines. There is significant potential for its widespread integration as a standard component of training practices to support staff. This integration would be especially beneficial for new staff members who have limited exposure to clinical environments. By providing realistic simulations and interactive learning experiences, VR training could play a crucial role in preventing serious incidents on wards and enhancing overall patient care.

Given the scarcity of research in this specific area, a systematic replication of our study could prove invaluable in testing the validity of our proposed aims and hypotheses. Systematic replications are generally more informative than literal or operational replications. They offer the advantage of uncovering new aspects and insights, as well as reassessing the primary objectives across varied subjects, measures, and other conditions without directly replicating the original study [37]. In future replications, prioritising a larger sample size will be crucial to ensure robust and generalisable results. Such an approach will not only validate the original proposal but also provide a more comprehensive understanding of any differences or consistencies across various subjects and measures.

CONCLUSION

To date, there has been no research exploring the effectiveness and feasibility of VR training in reducing RPs on inpatient

psychiatric wards. This study contributes to the growing body of literature assessing the efficacy of VR interventions in clinical settings. The results suggest that while VR is a promising tool for educational and therapeutic interventions, its impact on self-efficacy, anxiety, burnout, and compassion within a short-term application may be limited. Our study highlights the challenges associated with implementing such a programme, even though staff have rated its usability highly. With the advancement of VR technology, it's possible to create diverse scenarios and simulations tailored to various healthcare environments. This progress can empower staff with better training and preparation for handling a wide array of situations. Future research should focus on longitudinal studies with larger sample sizes and more diverse populations to fully ascertain the potential benefits of VR in reducing RPs and enhancing psychological well-being among clinical practitioners.

FOOTNOTES

Author contributions: Peter P innovated the project. Peter P and Delanerolle G conceptualised the project and wrote the first draft. Phiri P, Boulter I, and Delanerolle G developed the protocol and wrote the manuscript draft; Salmon J and Sajid S collected the data; Clarke J managed virtual reality administration; Phiri P, Liu Y, Yang X, Shi JQ, and Delanerolle G conducted the analysis; all authors critically appraised the manuscript; and all authors approved the final manuscript.

Institutional review board statement: The study received Health Research Authority (HRA) and Health and Care Research Wales (HCRW) approval. REC Reference: 22/HRA/3030. IRAS number: 317489.

Clinical trial registration statement: The manuscript was not applicable to apply for a clinical trial registration.

Informed consent statement: All study participants, or their legal guardian, provided informed written consent prior to study enrolment.

Conflict-of-interest statement: Phiri P has received research grant from Novo Nordisk, and other, educational from Queen Mary University of London, other from John Wiley & Sons, other from Otsuka, other from Janssen, outside the submitted work. All other authors report no conflict of interest. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or the Academic institutions.

Data sharing statement: The datasets generated during and/or analysed during the current study are not publicly available but are available from the corresponding author (Phiri P) on reasonable request.

CONSORT 2010 statement: The authors have read the CONSORT 2010 Statement, and the manuscript was prepared and revised according to the CONSORT 2010 Statement.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

Country of origin: United Kingdom

ORCID number: Peter Phiri 0000-0001-9950-3254; Joe Salmon 0009-0001-2456-2003; Sana Sajid 0000-0001-5737-7552; Jian-Qing Shi 0000-0002-2924-1137; Gayathri Delanerolle 0000-0002-9628-9245.

S-Editor: Wang JJ

L-Editor: A

P-Editor: Yu HG

REFERENCES

- 1 **Mooney M, Kanyeredzi A.** 'You get this conflict between you as a person and you in your role...that changes you': A thematic analysis of how inpatient psychiatric healthcare staff in the UK experience restraint, seclusion, and other restrictive practices. *Int J Ment Health Nurs* 2021; **30**: 1703-1712 [PMID: 34494346 DOI: 10.1111/inm.12926]
- 2 **Mind.** Mental health crisis care: physical restraint in crisis. [cited 15 January 2024]. Available from: https://www.mind.org.uk/media-a/4378/physical_restraint_final_web_version.pdf
- 3 **Huckshorn KA.** Reducing seclusion and restraint use in inpatient settings: a phenomenological study of state psychiatric hospital leader and staff experiences. *J Psychosoc Nurs Ment Health Serv* 2014; **52**: 40-47 [PMID: 25310675 DOI: 10.3928/02793695-20141006-01]
- 4 **Rakhmatullina M, Taub A, Jacob T.** Morbidity and mortality associated with the utilization of restraints: a review of literature. *Psychiatr Q* 2013; **84**: 499-512 [PMID: 23649219 DOI: 10.1007/s11126-013-9262-6]
- 5 **Azeem M, Aujla A, Rammerth M, Binsfeld G, Jones RB.** Effectiveness of six core strategies based on trauma informed care in reducing seclusions and restraints at a child and adolescent psychiatric hospital. *J Child Adolesc Psychiatr Nurs* 2017; **30**: 170-174 [PMID: 30129244 DOI: 10.1111/jcap.12190]
- 6 **Goulet MH, Larue C.** Post-Seclusion and/or Restraint Review in Psychiatry: A Scoping Review. *Arch Psychiatr Nurs* 2016; **30**: 120-128

- [PMID: 26804513 DOI: 10.1016/j.apnu.2015.09.001]
- 7 **Muir-Cochrane E**, O'Kane D, Oster C. Fear and blame in mental health nurses' accounts of restrictive practices: Implications for the elimination of seclusion and restraint. *Int J Ment Health Nurs* 2018; **27**: 1511-1521 [PMID: 29521451 DOI: 10.1111/inm.12451]
 - 8 **Pedersen ML**, Gildberg F, Baker J, Damsgaard JB, Tingleff EB. Ethnic disparities in the use of restrictive practices in adult mental health inpatient settings: a scoping review. *Soc Psychiatry Psychiatr Epidemiol* 2023; **58**: 505-522 [PMID: 36454269 DOI: 10.1007/s00127-022-02387-8]
 - 9 Violence and Aggression: Short-Term Management in Mental Health, Health and Community Settings: Updated edition. London: British Psychological Society (UK); 2015- [PMID: 26180871]
 - 10 Mental Health Problems in People with Learning Disabilities: Prevention, Assessment and Management. London: National Institute for Health and Care Excellence (NICE); 2016 Sep- [PMID: 27683921]
 - 11 **Stewart D**, Van der Merwe M, Bowers L, Simpson A, Jones J. A review of interventions to reduce mechanical restraint and seclusion among adult psychiatric inpatients. *Issues Ment Health Nurs* 2010; **31**: 413-424 [PMID: 20450344 DOI: 10.3109/01612840903484113]
 - 12 **12 Deveau R**, Leitch S. Implementation of policy regarding restrictive practices in England. *Tizard Learn Disabil Rev* 2020; **25**: 1-8 [DOI: 10.1108/TLDR-05-2018-0016]
 - 13 **NHS digital**. Restrictive intervention reporting in adult acute and adult learning disabilities inpatient services. [cited 15 January 2024]. Available from: <https://webarchive.nationalarchives.gov.uk/ukgwa/20230506043909/https://digital.nhs.uk/data-and-info>
 - 14 **Department of Health and Social Care**. Code of practice: Mental Health Act 1983. [cited 10 January 2024]. Available from: <https://www.gov.uk/government/publications/code-of-practice-mental-health-act-1983>
 - 15 **Care Quality Commission**. Mental Health Act. A focus on restrictive intervention reduction programmes in inpatient mental health services. [cited 10 January 2024]. Available from: https://www.cqc.org.uk/sites/default/files/201701207b_restrictivepractice_resource.pdf
 - 16 **Bell IH**, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice. *Dialogues Clin Neurosci* 2020; **22**: 169-177 [PMID: 32699517 DOI: 10.31887/DCNS.2020.22.2/ivalmaggia]
 - 17 **Tang YM**, Ng GWY, Chia NH, So EHK, Wu CH, Ip WH. Application of virtual reality (VR) technology for medical practitioners in type and screen (T&S) training. *J Comput Assist Learn* 2020; **37** [DOI: 10.1111/jcal.12494]
 - 18 **Hood RJ**, Maltby S, Keynes A, Kluge MG, Nalivaiko E, Ryan A, Cox M, Parsons MW, Paul CL, Garcia-Esperon C, Spratt NJ, Levi CR, Walker FR. Development and Pilot Implementation of TACTICS VR: A Virtual Reality-Based Stroke Management Workflow Training Application and Training Framework. *Front Neurol* 2021; **12**: 665808 [PMID: 34858305 DOI: 10.3389/fneur.2021.665808]
 - 19 **Mulla M**, Sharma D, Moghul M, Kailani O, Dockery J, Ayis S, Grange P. Learning basic laparoscopic skills: a randomized controlled study comparing box trainer, virtual reality simulator, and mental training. *J Surg Educ* 2012; **69**: 190-195 [PMID: 22365864 DOI: 10.1016/j.jsurg.2011.07.011]
 - 20 **Vaughan N**, Dubey VN, Wainwright TW, Middleton RG. A review of virtual reality based training simulators for orthopaedic surgery. *Med Eng Phys* 2016; **38**: 59-71 [PMID: 26751581 DOI: 10.1016/j.medengphy.2015.11.021]
 - 21 **Matthews T**, Tian F, Dolby T. Interaction design for paediatric emergency VR training. *Virtual Real Intell Hardw* 2020; **2**: 330-344 [DOI: 10.1016/j.vrih.2020.07.006]
 - 22 **Zhao J**, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. *BMC Med Educ* 2020; **20**: 127 [PMID: 32334594 DOI: 10.1186/s12909-020-1994-z]
 - 23 **Kyaw BM**, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, Divakar U, Masiello I, Kononowicz AA, Zary N, Tudor Car L. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J Med Internet Res* 2019; **21**: e12959 [PMID: 30668519 DOI: 10.2196/12959]
 - 24 Healthcare Simulation. Virti | Healthcare Simulation | HealthySimulation.com [Internet]. [cited 12 January 2024]. Available from: <https://www.healthysimulation.com/medical-simulation/vendors/virti/>
 - 25 **Young A**, Aquilina A. Use of Virtual Reality to Support Rapid Upskilling of Healthcare Professionals during COVID-19 Pandemic. In: Jung T, Dalton J. Management for Professionals. Cham: Springer, 2021
 - 26 **Schwarzer R**, Jerusalem M. General Self-Efficacy Scale (GSE). [cited 15 January 2024]. Available from: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F00393-000>
 - 27 **Iversen AC**, Kraft P, Røysamb E. Perceived self-efficacy in health behaviour research: Conceptualisation, measurement and correlates. *Psychol Health* 2000; **15**: 51-69 [DOI: 10.1080/08870440008400288]
 - 28 **Schwarzer R**, Mueller J, Greenglass E. Assessment of perceived general self-efficacy on the internet: Data collection in cyberspace. *Anxiety Stress Coping* 1999; **12**: 145-161 [DOI: 10.1080/10615809908248327]
 - 29 **Spitzer RL**, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166**: 1092-1097 [PMID: 16717171 DOI: 10.1001/archinte.166.10.1092]
 - 30 **Schaufeli WB**, Desart S, De Witte H. Burnout Assessment Tool (BAT)-Development, Validity, and Reliability. *Int J Environ Res Public Health* 2020; **17**
 - 31 **Williams DR**. Measuring discrimination resource. [cited 15 January 2024]. Available from: https://scholar.harvard.edu/files/davidrwilliams/files/measuring_discrimination_resource_june_2016.pdf
 - 32 **Gilbert P**, Catarino F, Duarte C, Matos M, Kolts R, Stubbs J, Ceresatto L, Duarte J, Pinto-Gouveia J, Basran J. The development of compassionate engagement and action scales for self and others. *J Compassionate Health Care* 2017; **4**: 4 [DOI: 10.1186/s40639-017-0033-3]
 - 33 **Bangor A**, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J Usability Stud* 2009; **4**: 114-123
 - 34 **Salaja B**, Feeley A, Feeley I, Sheehan E, Merghani K. Virtual reality simulation in orthopaedic surgical training during periods of restricted clinical hours: a systematic review. *J Surg Simul* 2021 [DOI: 10.1102/2051-7726.2021.0003]
 - 35 **Mazur LM**, Khasawneh A, Fenison C, Buchanan S, Kratzke IM, Adapa K, An SJ, Butler L, Zebrowski A, Chakravarthula P, Ra JH. A Novel Theory-Based Virtual Reality Training to Improve Patient Safety Culture in the Department of Surgery of a Large Academic Medical Center: Protocol for a Mixed Methods Study. *JMIR Res Protoc* 2022; **11**: e40445 [PMID: 36001370 DOI: 10.2196/40445]
 - 36 **de Winter JC**, Dodou D, Wieringa PA. Exploratory Factor Analysis With Small Sample Sizes. *Multivariate Behav Res* 2009; **44**: 147-181 [PMID: 26754265 DOI: 10.1080/00273170902794206]
 - 37 **Schmidt FL**, Oh I. The crisis of confidence in research findings in psychology: Is lack of replication the real problem? *Arch Sci Psychol* 2016; **4**: 32-37 [DOI: 10.1037/arc0000029]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: office@baishideng.com
Help Desk: <https://www.f6publishing.com/helpdesk>
<https://www.wjgnet.com>

