Systematic Review of Dependent Variables and Outcome Measures in XR Psychological Interventions with a Complementary Checklist

A THESIS DISSERTATION

Submitted to the Faculty of Behavioural, Management and Social Sciences (BMS)

in partial fulfilment of the requirements for the degree of

Bachelor of Science (B.Sc.) in Psychology

by

Paola Flores

University of Twente, Netherlands

July 2024

© [2024] by [Paola Flores]
Approval of the Thesis Dissertation

SYSTEMATIC REVIEW OF DEPENDENT VARIABLES AND OUTCOME MEASURES IN XR PSYCHOLOGICAL INTERVENTIONS WITH A COMPLEMENTARY CHECKLIST

This Thesis dissertation by Paola Flores has been approved by the supervisors below, who recommend it be accepted by The University of Twente in partial fulfillment of requirements for the degree of

Bachelor of Science (B.Sc.) in Psychology

Supervisors:

Funda Yildirim, Ph.D., Chair

Simone Borci, Ph.D.

Abstract

As extended reality (XR) technology advances, so does the exploration of its potential within psychological domains. However, its application in psychology is still novel, resulting in a lack of a widely accepted framework for developing effective XR psychological interventions, limiting comparability across studies. Establishing such a framework requires a thorough understanding of the current landscape of XR interventions. This systematic review provides a comprehensive overview of the current use of XR technology in psychological interventions, focusing on the psychological constructs studied and their measurement methods. A systematic search using Elsevier Scopus identified 167 studies that met the inclusion criteria. Seven overarching categories of dependent variables were identified, including mental health, exposure therapy, cognitive functioning, health, social well-being, and behavioural and attitudinal change. Examination of the measurement methods across these studies revealed 497 outcome measures, differentiated between intervention-related or participant-related measures. Participant-related measures were further classified as self-report, behavioural or physiological measures. Self-report measures were the most prevalent across all categories, while the use of behavioural and physiological measures was comparable. Inconsistencies in measuring similar dependent variables underscored the need for methodological standardisation, prompting the development of a checklist based on the review's findings. Ultimately, this systematic review provides an overview of the current dependent variables studied via XR and their corresponding measurement methods. Future research could further refine and broaden the checklist for broader applicability across psychological domains.

Keywords: Extended reality, psychological interventions, systematic review, dependent variables, outcome measures, methodological standardisation, checklist development.

Table of Contents

List of Figures 7
List of Tables 8
Introduction 9
Methodology 17
Overview 17
Search Strategy and Eligibility Criteria 17
Review Procedure 20
Data Extraction 23
Results 27
Literature Characteristics 27
Method of Classification 32
Categories Identified 35
Mental health 35
Exposure therapy 42
Mental Health and Cognitive Functioning 45
Cognitive Functioning 47
Health 51
Social Well-being 54
Behavioural and Attitudinal Change 58
Outcome Measures Identified 60
Checklist 76
Discussion 79
General Interpretation of Results 79
Evaluation of Dependent Variables Studied Using XR Technology 80
Evaluation of Outcome Measures Usage 82
Implications of the Results 86
Limitations of Evidence Included in the Review 90
Limitations of the Review Process 91
Conclusion 92
References 94

Appendix A. Article Distribution by Year of Publication……..……..104
Appendix B. Inclusion and Exclusion Keywords as Specified in Rayyan……...…105
Appendix C. Reasons for Article Exclusion in Rayyan…….……..106
Appendix D. Reasons for Excluding Articles in the Data Extraction Table……...107
Appendix E. Application Areas: Descriptions and Distribution Across Studies……108
E1. Descriptions……...109
E2. Distribution Across Studies for "All Studies" and "Included Studies".……111
Appendix F. Mental Health: Dependent Variables & Outcome Measures…….…….112
Appendix G. Exposure Therapy: Dependent Variables & Outcome Measures…….……..144
Appendix H. MH & CF: Dependent Variables & Outcome Measures……..……..148
Appendix I. Cognitive Functioning: Dependent Variables & Outcome Measures…….150
Appendix J. Health: Dependent Variables & Outcome Measures…….……..157
Appendix K. Social Well-being: Dependent Variables & Outcome Measures……...167
Appendix L. Behavioural and Attitudinal Change: Dependent Variables & Outcome Measures……...…181
Appendix M. Overview of Main Categories, Sub-Categories, and Further Classifications……189
Appendix N. Intervention-Related Questionnaires…….……192
Appendix O. Adverse Events……...195
Appendix P. Definitions of Terms found in Outcome Measures…….……196
Appendix Q. Participant-Related Questionnaires……....197
Q1. Self-Report Measures……197
Q2. Behavioural Measures……...…234
Q3. Physiological Measures……247
Appendix R. Checklist Mind Maps……...……..251

List of Figures

Figure 1. PRISMA Flowchart Detailing the Study Selection Process……..…..26

Figure 2A. Distribution of Application Areas for “Included Studies” (1/2)..……...…….…29

Figure 2B. Distribution of Application Areas for “Included Studies” (2/2)……..30

Figure 3A. Frequencies of Studies by Application Area for “All Studies” and “Included Studies” (1/2)……...……..31

Figure 3B. Frequencies of Studies by Application Area for “All Studies” and “Included Studies” (2/2)……...……..32

Figure 4. Distribution of Articles Across Main Categories Based on Dependent Variables……34

Figure 5. Distribution of Articles Across Subcategories Within Main Categories……35

Figure 6. Distribution of Self-Report Measures Across Categories……..62

Figure 7. Distribution of Self-Report Measures Across Categories…….….68

Figure 8. Distribution of Physiological Measures Across Main Categories……..72

List of Tables

Table 1. Distribution of Studies in "Mental Health" Category by Outcome Measure Type…..…43

Table 2. Distribution of Studies in "Exposure Therapy" Category by Outcome Measure Type...46

Table 3. Distribution of Studies in "Mental Health and Cognitive Functioning" Category by Outcome Measure Type…….48

Table 4. Distribution of Studies in "Cognitive Functioning" Category by Outcome Measure Type……...51

Table 5. Distribution of Studies in "Health" Category by Outcome Measure Type…….54

Table 6. Distribution of Studies in "Social Well-being" Category by Outcome Measure Type……...……58

Table 7. Distribution of Studies in "Behavioural and Attitudinal Change" Category by Outcome Measure Type…….60

Table 8. Classification of Self-Report Measures into Categories and Subcategories and the Total Number of Questionnaires Identified for Each Psychological Construct……..63

Table 9. Classification of Behavioural Measures into Categories and Subcategories……...69

Table 10. Classification of Physiological Measures into Categories and Frequencies……..73

Introduction

The World Health Organisation (WHO) identifies mental health challenges, especially anxiety and depression, as the primary contributor to years lived with disability, resulting from inadequate treatment provision. Barriers such as limited capacity, inaccessibility, high costs, and stigma often prevent individuals with mental health conditions from accessing necessary services (WHO, 2024). Traditionally, psychological interventions have been administered face-to-face by trained professionals, however, novel approaches to delivering these are now being explored to address these barriers.

Psychological interventions comprise therapeutic techniques grounded in psychological theory, designed to enhance cognitive functioning, psychological well-being, and address behavioural issues (National Collaborating Centre for Mental Health (UK), 2011). To effectively determine the efficacy of these interventions, researchers must monitor changes in targeted psychological constructs, including alterations in emotion, motivation, attention, arousal, and cognition (Diamond & Otter-Henderson, 2009, pp. 370–371). Psychological constructs are characterised by their inability to be assessed through observation, often representing tendencies to think, feel or behave across a variety of situations or internal processes, such as nervous system activation (Price et al., n.d.). Additionally, psychological interventions target other constructs including well-being (e.g. quality of life), personality traits (e.g. openness), cognitive functioning (e.g. memory) emotional states (e.g. disgust), attitudes (e.g. toward climate change), and abilities (e.g. time management) (Ryser, V.-A, 2023; Price et al., n.d.).

Psychometric measurement, the process of evaluating psychological constructs, involves a systematic procedure assigning scores to individuals to represent the variable of interest according to its operational definition (Price et al., n.d.). These measurements are divided into three broad

categories: self-report measures, behavioural measures, and physiological measures. Self-report measures involve participants providing insights into their own cognition, motivation, behaviour, emotion, or physical state, allowing for a differentiated assessment of human thought compared to other psychometric methods (National Collaborating Centre for Mental Health (UK), 2011). However, self-report is "limited to the assessment of processes that are accessible to consciousness" and language, and it is susceptible to memory bias, potentially compromising the objectivity of results (National Collaborating Centre for Mental Health (UK), 2011). Behavioural measures regard the observation and notation of a specific aspect of a participant's behaviour by an impartial observer, either in a structured laboratory environment or in a natural environment (Price et al., n.d.). Lastly, physiological measures consider the recording of any physiological process using a variety of instruments across laboratory or natural settings (Gaffey & Wirth, 2014). Psychological interventions often employ multiple methods to define and measure a particular construct, a practice known as converging operations, to increase the validity of results (Price et al., n.d.). Consistent patterns across the results of these psychometric approaches often indicate strong evidence that the construct is being effectively measured (Price et al., n.d.).

Psychophysiological measures evaluate the relationship between psychological processes and bodily states, providing a holistic understanding of their influence in behaviour. Additionally, they allow continuous data collection, enabling pattern detection amidst the effects of ongoing exposure to stimuli (Gaffey & Wirth, 2014). Among the most frequently used psychophysiological measures are those that detect variations in physiological arousal stemming from both the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) (Gaffey & Wirth, 2014). These measures encompass electrodermal activity, blood pressure, heart rate, heart rate variability, and respiration (Gaffey & Wirth, 2014).

Electrodermal activity (EDA) is a noninvasive measure that records the skin's electrical activity by placing sensors to detect changes in sympathetic nervous system (SNS) activation. It relies on the activation of the eccrine glands and does not necessarily require sweat secretion (Gaffey & Wirth, 2014)

Moreover, blood pressure (BP) regards assessing the arterial pressure of circulating blood, typically using ausculatory or oscillometric methods. Ausculatory measurement involves placing a cuff on the upper arm along with a stethoscope, whereas oscillometric methods measure pressure fluctuations with a BP cuff. BP levels correlate with multiple psychological factors and tend to increase in response to stress, negative emotions, and SNS arousal (Gaffey & Wirth, 2014).

Measuring heart rate (HR) involves recording the time interval between successive heartbeats, commonly expressed in beats per minute. Previous literature suggests that emotionally stimulating experiences, spanning from excitement to fear, result in heightened heart rate responses, particularly influenced by the amygdala’s assessment of perceived stimuli, responding to emotion by SNS activation (Gaffey & Wirth, 2014).

Furthermore, heart rate variability (HRV) is employed to represent PNS cardiac control focusing on variations in vagal sinoatrial activity, particularly measured by evaluating fluctuations in the beat-to-beat intervals. This measurement is derived from a combination of other physiological measurements, including BP, electrocardiography (ECG) and a pulse wave signal. HRV is recognized for its important role in emotion regulation and coping mechanisms, with higher HRV indicative of a greater capacity for effective emotion regulation. In addition, it serves as a commonly utilised index for anxiety, as lower HRV is often observed in individuals with an underlying anxiety disorder compared to control groups (Gaffey & Wirth, 2014).

Ultimately, respiration measures regard the recording of the “frequency and amplitude of breathing” using strain gauges positioned around the individual to capture both upper and lower respiratory movements. These gauges, comparable to belts, expand and contract with the individual’s inhalations and exhalations. Studies indicate that emotional and psychological responses typically triggering SNS activity result in heightened breathing rates (Gaffey & Wirth, 2014).

Given that psychophysiological measures rely on increases in sympathetic nervous system (SNS) arousal stemming from emotionally charged stimuli, traditional psychological interventions may face challenges in fully immersing participants in such environments to evoke authentic reactions. Consequently, there is an increasing acknowledgement of Extended Reality (XR) technology's potential to improve these interventions, emphasising the need to evaluate its current application. XR encompasses several immersive technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). In VR, the user is deprived of the visual and auditory input from the physical environment by a head-mounted display (HMD) and headphones, thus, enabling the experience of a completely virtual environment (Lee, 2020). Conversely, in AR, the user engages with digital objects superimposed onto the physical environment, commonly perceived through specially designed glasses or the screen of a mobile phone (Lee, 2020). MR combines elements of both VR and SR to enable a blend of digital and real-world elements (Lee, 2020). Comparably, all three XR technologies allow for the possibility of immersive and controlled settings for interventions.

XR technologies offer psychological interventions the opportunity to provide users with immersive experiences. While there is no standard definition of immersion, literature widely regards it as a multidimensional construct (Lee, 2020). Among others, three distinct dimensions of

immersion are highlighted: passive immersion, active immersion, and challenge-based immersion (Lee, 2020). Passive immersion is induced by the sensory stimulation of the virtual environment (VE) and is often connotated with the sense of 'presence' (Lee, 2020). Active immersion is obtained by the involvement of the user in the narrative of the content presented (Lee, 2020). Lastly, challenge-based immersion is experienced when the user participates in a challenging task or activity within the XR experience (Lee, 2020). Using these dimensions of immersion, psychological interventions can simulate real-life scenarios, provide interactive experiences, and deliver biofeedback.

XR-generated simulations of real-life scenarios enable users to interact with the XR environment while receiving real time feedback from it. This responsiveness to spontaneous user input enhances the perception of realism within the simulated environment and can be used to trigger and measure genuine emotional, psychological and physical responses in users through biofeedback (Geraets et al., 2021). Given these capabilities, early applications of computer-generates simulations emphasised their use for VR exposure therapy (VRET) interventions. Further developments highlighted the potential of XR to situate users in situations beyond real-life experiences, enabling the application of novel strategies in therapeutic contexts and broadening its use to investigate and treat mental disorders (Geraets et al., 2021).

According to Geraets et al. (2021), immersive XR is primarily used in mental health intervention studies targeting anxiety disorders, psychotic disorders, substance-related disorders, eating disorders, and depression. Despite the broad scope of disorders targeted, they identify anxiety as the "most mature field" studied through VR, encompassing methods such as VRET and VR-based cognitive behavioural therapies (VR-CBTs). VRET interventions have evolved beyond their initial focus on specific phobias to encompass a wider array of anxiety-related disorders

including social anxiety disorder, panic disorder, and post-traumatic stress disorder (Geraets et al., 2021). While the success rates for VRET interventions surpass those of waiting-list or placebo conditions, they remain comparable to conventional treatments (Geraets et al., 2021).

The exploration of immersive XR's potential across various psychological domains has spurred innovative developments in the field, including automated VR treatment, embodiment techniques, and perspective-taking techniques. Automated VR treatments involves standalone or self-guided approaches aimed at making psychological therapy more readily accessible and provide a complementary tool to face-to-face treatment (Geraets et al., 2021). Developments in embodiment techniques focus on instilling feelings of body ownership or a body-swapping illusion in users, where they perceive the virtual body as their own, strengthened by synchronous tactile stimulation and synchronous movements of both bodies (Geraets et al., 2021). Applications of this novel technique enable changes in attitudes, cognitions, self-perception, and behavioural patterns. The last development mentioned by Geraets et al. (2021) involves perspective-taking techniques, allowing users to role-play with avatars and later observe the conversation as an outsider. This ability to experience the same virtual scenario from multiple perspectives enhances understanding of behaviour and interactions, particularly when targeting constructs like self-criticism, helping intentions, or empathy. However, despite the developments and growing interest in the possibilities brought about by XR for applications in psychological interventions, technological literacy, accessibility, and acceptance of XR technologies remains low among end users and professionals, leading to limited use of these technologies (Antanavičius, 2023).

The limited application of XR to psychological interventions as a tool for enhancing these has led to the lack of a scholarly accepted structured framework for the development of effective psychological interventions applicable across XR applications. This gap highlights the need for

the development of a standardised methodology to facilitate the systematic analysis and comparison of diverse XR environments utilised in psychological interventions. Such methodology should focus on establishing guidelines or protocols for the consistent design, implementation and evaluation of XR environments, ultimately improving reliability and reproducibility across studies.

Prior to addressing these shortcomings, a holistic understanding of the current use of XR psychological interventions is needed. To achieve this, a structured and unbiased approach to collect data is required, which involves conducting a systematic literature review. A systematic review comprises a detailed plan and search strategy formulated from a research question, aimed at identifying, evaluating, and synthesising all relevant studies on a given topic (Uman, 2011). This type of literature review minimises selection bias through its systematic and reproducible approach to searching, selecting, and assessing studies according to the predefined selection criteria (Uman, 2011).

The goals of this systematic literature review are twofold. Firstly, a comprehensive overview of the existing literature concerning XR-based psychological interventions involving self-report, behavioural or physiological outcome measures, organising the studies according to their dependent variable will be conducted. Secondly, at the end of the review process as an additional goal, a checklist will be developed based on the dependent variables identified in the literature. This repository of items aims to provide researchers with points to consider when developing their own XR psychological interventions based on their dependent variable.

Upon presenting the methodology, the results section is divided into five subsections, including a synopsis of the characteristics of the selected literature, the method of classification employed along with the identified categories, data on the outcome measures utilised across the

literature, and an overview of the checklist proposed. Moreover, the discussion section includes a general interpretation of the findings, followed by an evaluation of patterns found across the dependent variables and outcome measures. Implications of the results and identified limitations are also presented in this section. Finally, the concluding section provides a synopsis of the key takeaways, emphasising the research question, most relevant results and their importance to the field.

Given the focus on exploring current XR psychological interventions, specifically regarding the variables studied and their methods of measurement, an exploratory research question was formulated to guide this literature review: Which psychological constructs are currently being studied in XR psychological interventions, and how are these constructs being measured? Aligned with the secondary goal of this thesis, another research question was determined: Can a checklist be developed to guide researchers in the design process of future XR psychological interventions, based on the categories of dependent variables identified in the systematic literature review?

Methodology

Overview

To gain a comprehensive understanding of the current landscape of XR intervention design, this study adopted a systematic methodological approach. The initial steps involved the formulation of research questions, definition of keywords and search terms, refinement of selection criteria, selection of a search engine, and conducting the search, extraction, and selection of data through a multi-level process. In the final paper selection phase, adherence to specific quality criteria within the studies, including the research question and the independent and dependent variables studied, was ensured. The study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis; Moher, 2009) framework to report findings (Pratama et al., 2022). PRISMA is a well-established guide tailored to improve the consistency of systematic literature reviews and meta-analyses, utilizing a checklist and a flowchart (Moher et al., 2009).

Search Strategy and Eligibility Criteria

Prior to the systematic search, a general consensus to the determine the scope of the literature review was conducted between the reviewers. It was decided that the literature would focus exclusively on psychological interventions utilising extended reality as a means of delivery. An intervention was considered psychological when there was application of psychological techniques and theories to promote prosocial or positive change in an individual. The dependent variable did not necessarily have to be a psychological construct for the study to qualify as a psychological intervention, provided it met the aforementioned criteria.

Given the broad applications of psychology across various domains, the potential overlap between psychological interventions and other similar types, particularly medical interventions, had to be taken into account in order to accurately discern and exclude these. Interventions were

considered medical when they involved the use of medical treatments, medications, or procedures to treat or prevent physical health conditions, including diseases, injuries and physiological dysfunctions (Smith et al., 2015).

Often, medical treatment approaches may integrate both medical and psychological interventions to treat the patient holistically; for example treating the physical complaints of chronic pain with medication while tending to the patient's distress with cognitive behavioural therapy (CBT). In these specific cases, it was decided that if the interventions focused primarily on treating the physical symptoms of a specific medical population, it would be excluded, regardless of whether psychological constructs were also assessed. Conversely, studies were included if the main objective of the intervention was to mitigate psychological distress associated with a medical condition.

Further considerations involved searching for psychological interventions that used extended reality with immersive technology. There was no preference for the type of XR used—whether VR, AR, or MR—as long as it met the immersion criteria. Given the novelty of immersive technology, a range for the publication years of the literature was not specified. Additionally, it was decided that only articles describing interventions would be included.

Elsevier Scopus was utilised as the sole search engine to gather scholarly literature. This decision was based on the extensive quantity of relevant and high-quality studies available from Scopus. Moreover, given the practical time constraints, it was preferred to conduct an in-depth review on this numerous sample of literature rather than a broader, less meticulous review that included more studies. The search procedure carried out in Scopus is detailed below.

To align with the systematic review's research question, the initial keywords specified in the advanced search query included 'psychology', 'virtual reality' and 'intervention'. The decision

to use the keyword ‘virtual reality’ instead of ‘extended reality’ was based on the lesser adoption of the term extended reality, as it is an umbrella term. It was noted that studies using augmented reality or mixed reality often included the term ‘virtual reality’ in their study’s keywords. Thus, using ‘virtual reality’ as a search term did not limit the studies to virtual reality alone but still encompassed all extended reality technologies. Additionally, the search parameters were adjusted to exclude articles with the terms ‘review’, ‘medical’, ‘critique’ and ‘meta-analysis’ in the title. The aforementioned parameters resulted in a total of 776 documents found.

Search results with the aforementioned search terms were determined to be overly medical, prompting the inclusion of additional exclusion parameters in the advanced search. The revised parameters were applied to article titles, abstracts, and keywords, resulting in the exclusion of medical terms including orthopaedic, cancer, cardiovascular, stroke, cardiac, COVID, respiratory, spinal, birth, tumour, pain, and management. Upon filtering out medical articles associated with the aforementioned terms, a total of 614 articles remained.

Aligned with the considerations mentioned above, parameters regarding the year of publication were not included in the search criteria. The search results yielded articles spanning from 1997 to 2024. The majority of articles fell within the range of 2017-2023, with the highest number published in 2023 (n= 98) and 2022 (n= 86). Refer to Appendix A for an exhaustive list of publication years across the retrieved articles.

After integrating the search terms and the exclusion parameters, the final search query inputted into Scopus was the following: Article title, Abstract, Keywords // psychology OR psychological; virtual AND reality; intervention; NOT meta AND analysis OR review OR critique; NOT medical OR orthopaedic OR cancer OR cardiovascular OR stroke OR cardiac OR

covid OR respiratory OR spinal OR birth OR tumour; OR pain AND management. All terms following the connector ‘NOT’ are exclusion parameters.

Review Procedure

After refining the search query in Scopus, the 614 remaining documents, along with their respective abstracts, were imported directly to the software application Rayyan. Rayyan is an AI-powered cloud-based platform tailored to researchers conducting systematic literature reviews and meta-analyses, that facilitate efficient management, screening, and analysis of the uploaded literature (Rayyan Help Center, 2024). Before screening the literature, the auto-resolver feature for duplicate articles was applied. Following this, keywords for inclusion and exclusion were specified, allowing the intelligent platform to automatically identify and highlight these terms across the abstracts. A comprehensive description of these keywords can be found in Appendix B. Furthermore, all articles were manually reviewed and classified as ‘Include’ (n= 275), ‘Exclude’ (n= 223), ‘Maybe’ (n= 112), or ‘Conflict’ (n= 2) based on the title and abstract.

Included Literature

Literature categorised as ‘include’ (n= 275) underwent further labelling to allow for the identification of prevalent themes. Employing an emergent coding approach, these labels were determined organically as the literature was reviewed, primarily encompassing specific populations and targeted disorders or diseases addressed by the interventions. Labels identified regarding disorders included autism spectrum disorder (ASD) (n=34), schizophrenia (n=15), seasonal affective disorder (SAD) (n=10), body dysmorphic disorder (BDD) (n=7), dementia (n=7), substance use disorder (SUD) (n=5), depressive disorder (n=5), cerebral palsy (n=2), Attention-deficit/hyperactivity disorder (ADHD) (n=2), borderline personality disorder (BPD) (n=1), and obsessive-compulsive disorder (OCD) (n=1). Three additional labels were developed,

aside from the ones described above, to categorise literature that served a specific purpose or employed particular study designs, denoted as ‘background theory’(n=68), ‘case study’ (n=7), or ‘qualitative study’ (n=2). Literature labelled as ‘background theory’ and ‘qualitative study’ was not considered for the results of this literature review, but were kept as potential resources for theoretical background.

Excluded Literature

To begin the screening process, article titles and availability of the paper were examined. Article titles utilizing keywords such as meta-analysis, literature review and critique, were excluded. Moreover, the link to view the full text at the publishing journal was tested, if the paper was not publicly available, only available in a foreign language, or did not have a DOI number, they were excluded. A total of 37 papers were excluded during the screening process.

After the screening process, additional excluded literature was annotated with a ‘reason for exclusion’ label for justification purposes. Exclusion reasons regarding the scope of the studies included those with a medical approach to the intervention, non-psychological interventions, and case studies. Evaluations of the current state of XR in psychological research, assessments of interventions, proposals for future XR research, summaries of psychological research, conference proceedings, research for future XR psychological interventions, position papers, and editorials were also excluded based on the article’s content and direction. Moreover, literature was excluded for the wrongful use of extended reality encompassing the following reasons: no use of immersive extended reality technology, use of XR but for research and not for delivery purposes, use of XR only in combination with other treatments or intervention types, and use of XR for cognitive assessment or diagnosis. Lastly, literature was excluded based on technical issues such as the absence of an abstract, lack of public availability of the article, availability only in a foreign

language, or duplicated articles. A comprehensive list of the exclusion reasons with their respective frequencies can be found in Appendix C.

Resolution of Conflicts Between Reviewers

After reading the abstracts, a total of 112 papers were categorised as 'maybe' and were accompanied by a comment explaining the reasons for uncertainty. These reasons included doubts about whether the study incorporated an intervention or was solely an investigation using XR, whether the technology used for the intervention was encompassed by the XR umbrella (e.g., "internet-based intervention"), and questions about whether the intervention was medical or psychological (e.g., a study on chronic migraines with some psychological variables measured). Moreover, there was uncertainty about whether studies that only included behavioural and/or physiological output measures, without psychological constructs, should be included. Some studies were categorised as ‘maybe’ if the abstract did not clearly indicate the what the intervention encompassed and the methodology used. One study was labeled as 'maybe' because it was unclear to the reviewer whether the use of hypnosis to reduce anxiety constituted a psychological intervention. Confusion also arose about whether studies that combined different intervention methods for a single group (e.g., a combination of VR exposure therapy and traditional exposure therapy) should be included, as the results of the VR experience could not be isolated from the non-VR intervention.

Conflicting literature was automatically categorised as such when two or more independent reviewers disagreed on whether to ‘include’ or ‘exclude’ the given paper. In this literature review, two reviewers were tasked with screening the studies, each reviewing 307 papers. Consequently, only two studies were identified as conflicting.

To resolve these uncertainties or disagreements, a third independent reviewer was tasked to appraise all studies marked as such. As a result of this review process, 14 articles were considered relevant to the literature review, while the remaining 91 were deemed unsuitable. Combining these results with the previously categorised literature, a total of 289 articles were classified as included, while 314 were classified as excluded.

Data Extraction

To facilitate the further screening process for the articles classified as 'included', all 289 articles were manually entered into a Microsoft Excel spreadsheet to allow for detailed data extraction. The data extraction table comprised several columns, including Author name, Keywords (from Rayyan), Article title, Application area, Research question, Procedure, Stimuli, Participants, Findings, Type of VR used, Citation, and Notes. To enhance efficiency, the 289 articles were split between the two reviewers. Articles designated solely for background information in the introduction were colour-coded dark blue (n=24), whereas those with missing information or not meeting the inclusion criteria were colour-coded red (n=98). It is important to note that due to the different scopes of the reviewers literature reviews, each reviewer applied distinct inclusion criteria, resulting in varying numbers of red color-coded articles and labels for exclusion. The excluded articles and the reasons for their exclusion are indicated in the data extraction table for each reviewer, an overview of this data can also be found in Appendix D. The remaining articles, identified as relevant to the literature review, were colour-coded as green and labeled with their corresponding category name (n=167). Colour coding was also applied to the 'Application area' variable to aid in visualising patterns, resulting in the emergence of 21 distinct categories. The 'Application area' categories, a description, and their corresponding prevalence are indicated in Appendix E. The 'Application area' labels were considered during

the development of the final categories but did not ultimately form the basis for the literature review. Moreover, five of the excluded studies were not allocated to an ‘Application area’ category as they did not fit suitably into any existing category.

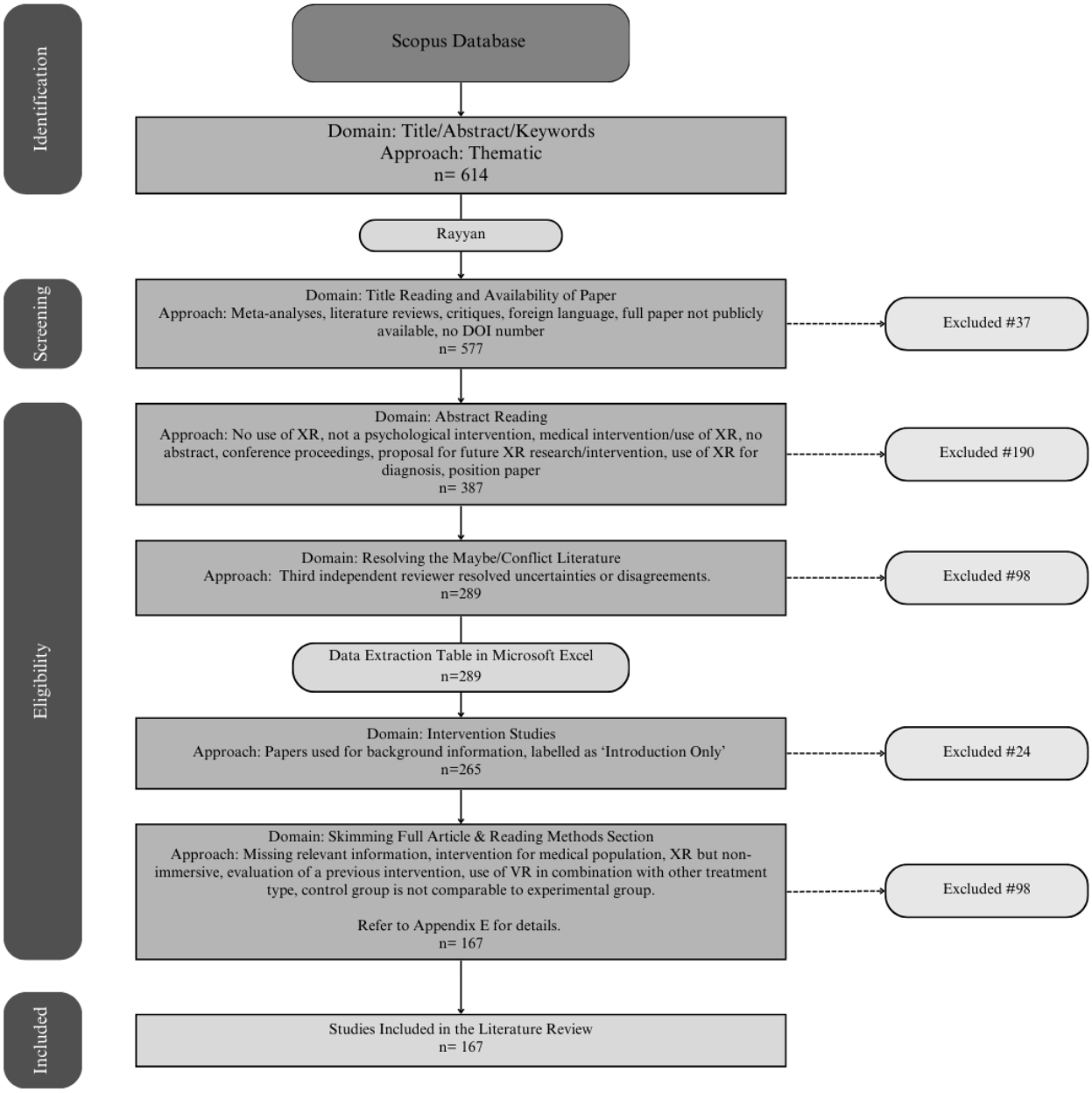
Consistent with each reviewer’s inclusion criteria, columns were added to the shared data extraction table for their respective articles. Four variables were specifically added for this literature review: Dependent variable, Self-report measures, Behavioural measures, and Physiological measures. These new variables are essential to this literature review as it aims to identify the dependent variables studied and how they are measured in existing literature on XR-based psychological interventions.

All included articles were thoroughly reviewed, with special emphasis on the Methods and Materials sections. The measures described in the interventions were classified into one of three outcome measure types: self-report, behavioural, or physiological. All measures utilized in each of the psychological interventions, including screening measures and self-designed questionnaires, were named in the data extraction table alongside their corresponding abbreviations. Furthermore, an exhaustive list of all outcome measures used across the included articles was compiled with a brief description of the constructs being assessed by each particular questionnaire or task. The outcome measures were ordered alphabetically, as well as categorised into 'Intervention-related measures' and 'Participant-related measures', where the latter was further categorised by outcome type.

Ultimately, 167 studies were included in this literature review. The PRISMA flowchart, as shown in Figure 1, illustrates the process used to exclude and finalize the array of studies for use, starting from the search engine results to the final selection.

Figure 1

PRISMA Flowchart Detailing the Study Selection Process



Note. The PRISMA flowchart above visualises the process used to screen and select articles for the literature review. The process is divided into four sections: "Identification", "Screening",

“Eligibility”, and “Included”. The “Screening” section involves the exclusion of articles based on superficial characteristics, such as genre or availability. Furthermore, the “Eligibility” section further examines the content and context of the studies, focusing on the inclusion criteria. The middle column depicts a step-by-step filtering of the studies, indicating the actions taken (“Domain”), the characteristics evaluated (“Approach”), and the number of remaining articles after applying these criteria. Finally, the rightmost column provides a count of excluded articles at each step.

Results

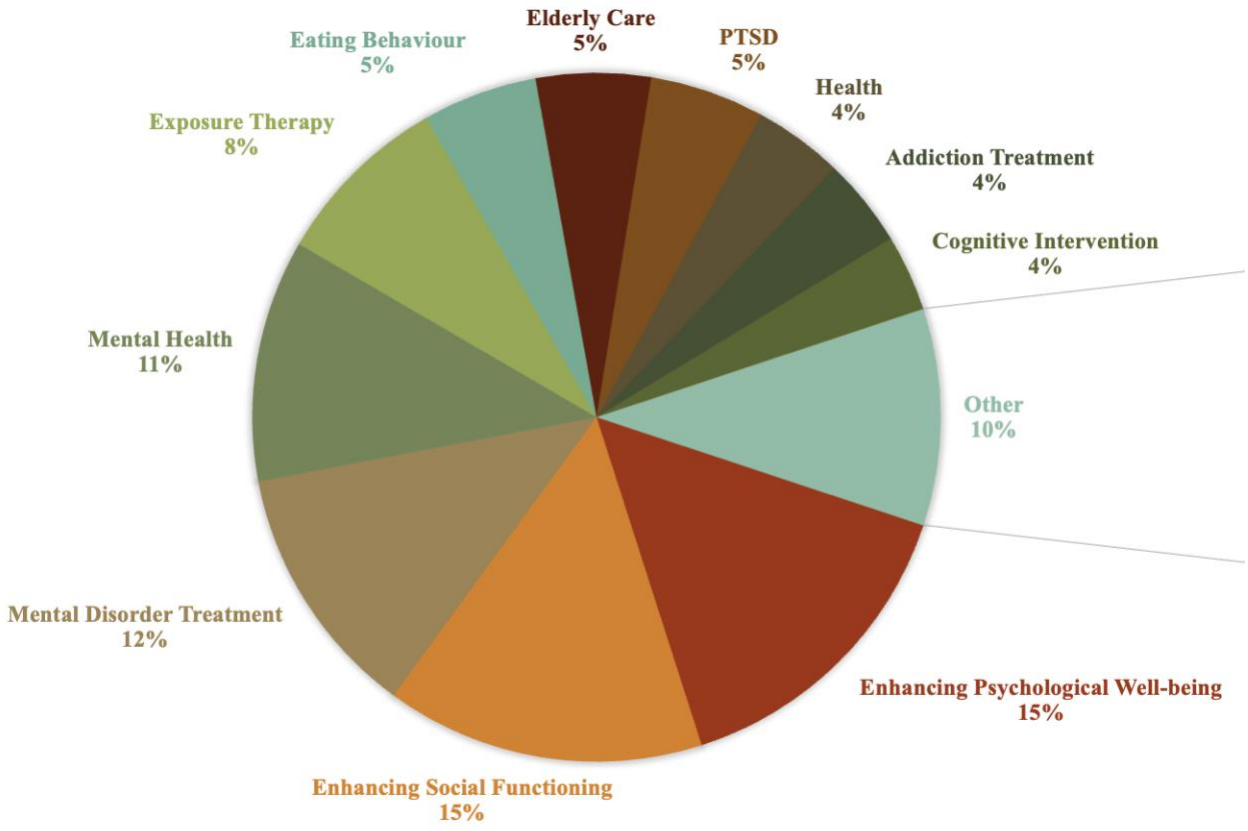
Literature Characteristics

To describe the final dataset, frequencies were calculated to illustrate the distribution of articles across years of publication and application areas. Regarding years of publication, half of the included articles (n= 83) correspond to the four most recent years, specifically 2020-2024. From the remaining 83 articles, 38 pertained to the years 2017-2019. Moreover, only five studies included in this literature review were published before 2010, across the years of 2009 (n=1), 2006 (n=1), 2002 (n=1), and 2000 (n=2) respectively.

Similarly, approximately half of the articles fall into four application areas (n=89): enhancing psychological well-being, enhancing social functioning, mental disorder treatment, and mental health (Figures 2A and 2B). The remaining 18 application areas have varied frequencies of articles, ranging from 1 to 9 articles per category. It is important to note that these application areas were identified based on the context in which the intervention is most useful, rather than the specific dependent variable or final category being studied, and thus, they are not directly comparable. Figures 3A and 3B illustrate the distribution of included studies across the application areas in comparison to the initial number of articles categorised by application area. For specific counts, refer to Appendix A for years of publication and Appendix D2 for application areas.

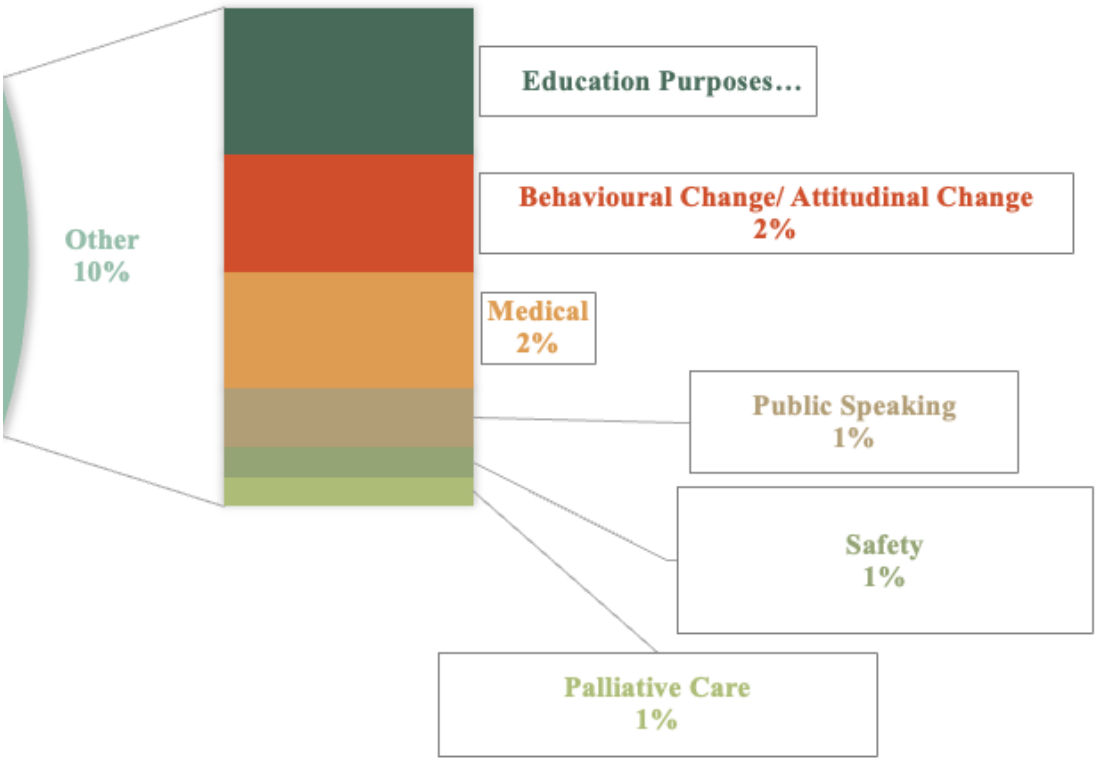
Figure 2

2A. Distribution of Application Areas for "Included Studies" (1/2)



Note. The piechart shows the distribution of studies across the application areas for the "Included Articles" exclusively, along with their respective percentages. The sector denoted as "Other" comprises the categories with the lowest number of studies, refer to Figure 2B for a detailed view of these categories.

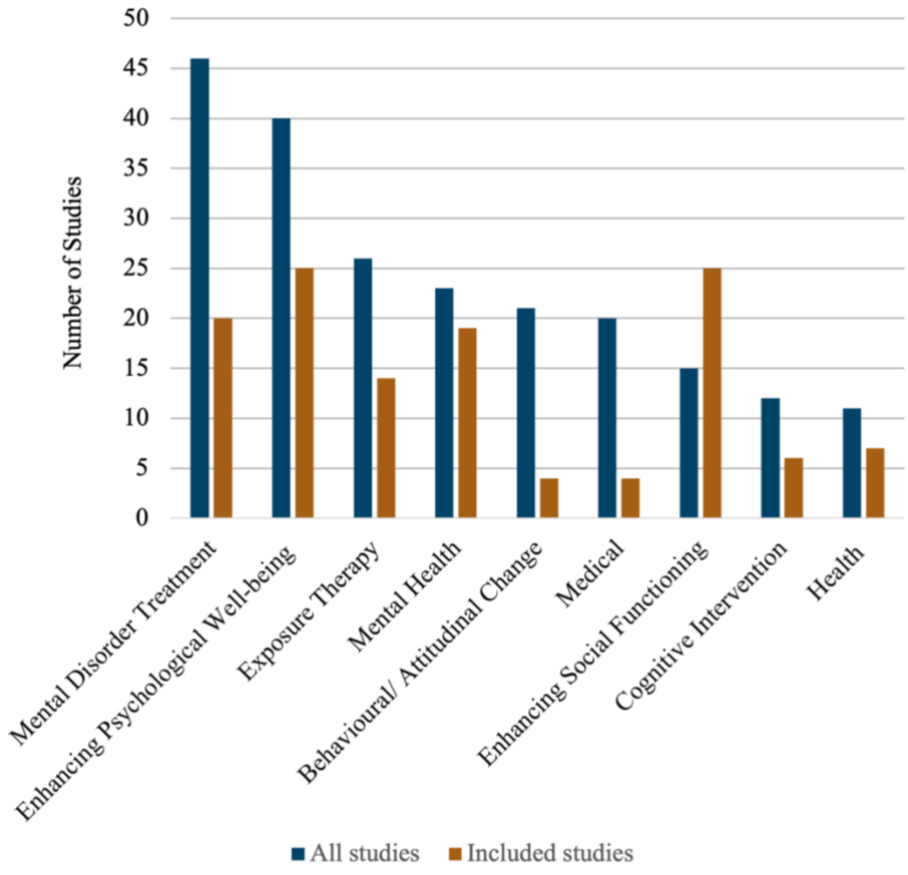
2B. *Distribution of Application Areas for "Included Studies" (2/2)*



Note. The segmented bar above shows the distribution of studies across the application areas denoted as "Other" in Figure 2A, with their respective percentages.

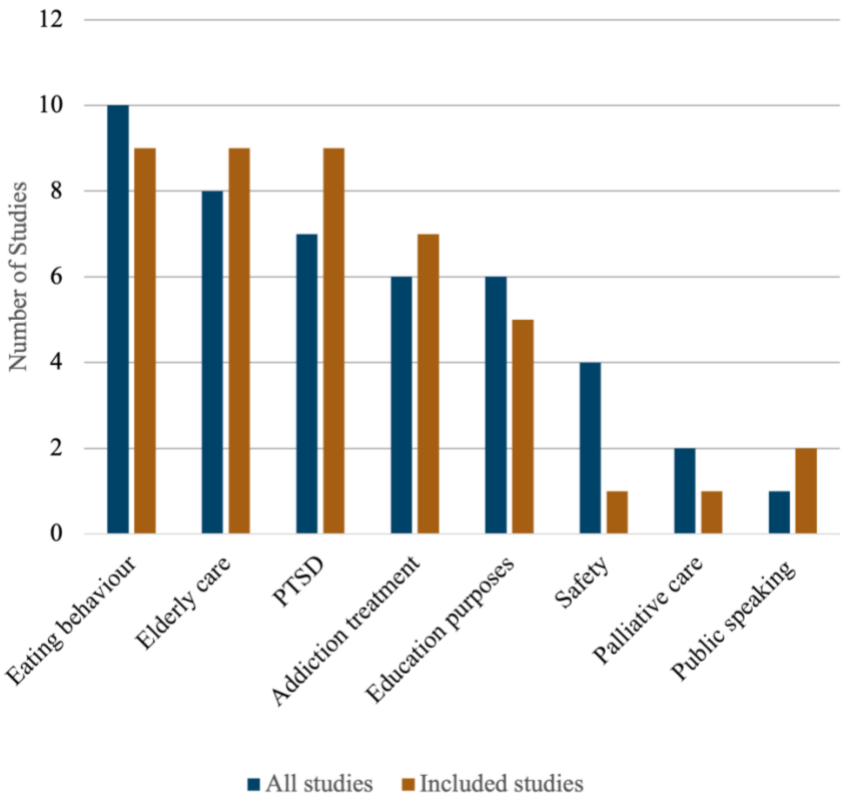
Figure 3

3A. Frequencies of Studies by Application Area for "All Studies" and "Included Studies" (1/2)



Note. The bar graph above illustrates the distribution of articles across the "Application Area" categories with the highest number of studies for "All Studies" and "Included Studies". For the remaining categories, refer to Figure 3B. As indicated in the PRISMA flowchart, the category "All Studies" encompassess all articles that were included in the data extraction table (n=289), while the "Included Studies" refer to the final selection of 167 articles. The figure aims to show the change in distribution proportions before and after the exclusion of data.

3B. Frequencies of Studies by Application Area for "All Studies" and "Included Studies" (2/2)



Note. The bar graph above illustrates the distribution of articles across the "Application Area" categories with the lowest number of studies for "All Studies" and "Included Studies". As indicated in the PRISMA flowchart, the category "All Studies" encompassess all articles that were included in the data extraction table (n=289), while the "Included Studies" refer to the final selection of 167 articles. The figure aims to show the change in distribution proportions before and after the exclusion of data.

Other noteworthy characteristics of the data include the absence of results from four studies due to ongoing data collection at the time of publication. Given the scope of this literature review on the dependent variable and its measurement, the absence of these results does not undermine

the validity of using the study's methodology as a data source. As mentioned previously, studies tailored to specific medical populations were excluded; however, interventions targeting populations with mental disorders were included.

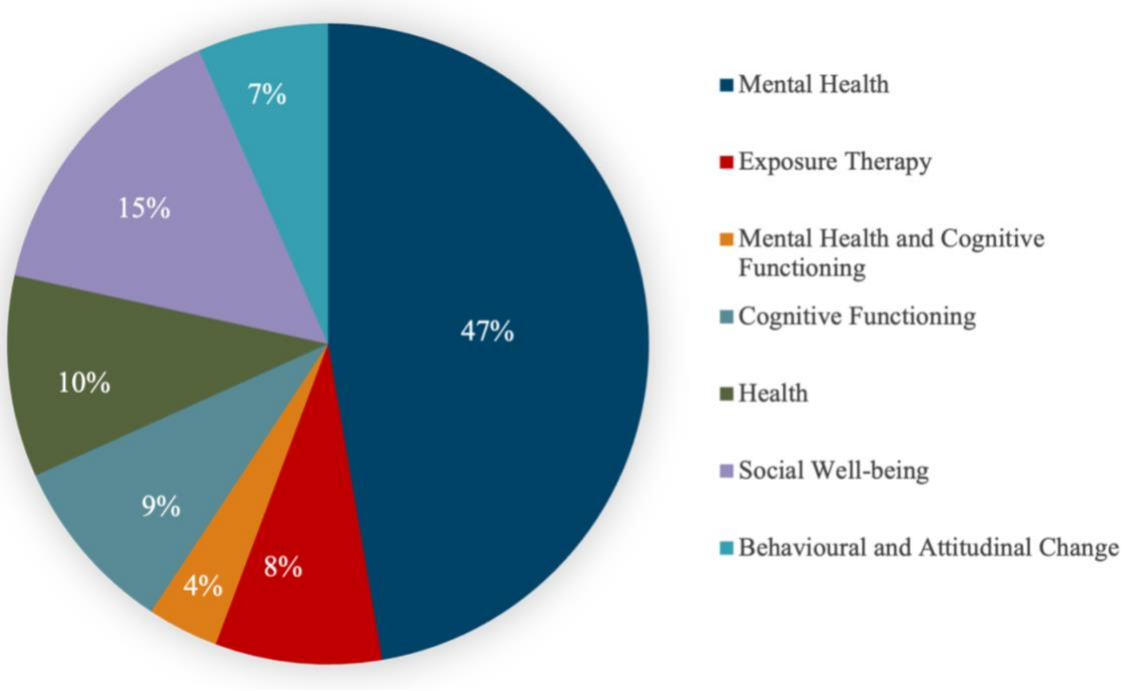
Again, a medical intervention was defined as an intervention targeting a population with a condition or disease that affects the body's physiological or biological systems, involving a diagnosis through clinical evaluation. Furthermore, psychological interventions tailored to populations with mental disorders were based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). According to the DSM-5, the included interventions encompassed disorders under the categories of "Schizophrenia Spectrum and Other Psychotic Disorders," "Neurodevelopmental Disorders," "Trauma- and Stressor-Related Disorders," "Depressive Disorders," and "Anxiety Disorders."

Method of Classification

After initially categorising the articles by area of application, patterns in the dependent variables within each area were identified. These patterns led to adjustments in the categories, including the removal of some categories and the development of new ones. The classification method for this literature review was based on the dependent variables studied in the XR psychological interventions. Patterns within each final category where further divided into subcategories. Seven main categories were identified within the retrieved literature: "Mental Health", "Exposure Therapy", "Mental Health and Cognitive Functioning", "Cognitive Functioning", "Health", "Social Well-being", and "Behavioural and Attitudinal Change". Refer to Figure 4 for an illustration of the distribution of studies across these categories, and Figure 5 for a more detailed depiction presenting the identified subcategories.

Figure 4

Distribution of Articles Across Main Categories Based on Dependent Variables



Note. Number of studies under "Mental Health" = 79, number of studies under "Exposure Therapy" = 14, number of studies under "Mental Health and Cognitive Functioning" = 6, number of studies under "Cognitive Functioning" = 15, number of studies under "Health" = 17, number of studies under "Social Well-being" = 25, number of studies under "Behavioural and Attitudinal Change" = 11, total number of studies = 167.

Figure 5

Distribution of Articles Across Subcategories Within Main Categories

Note. Subcategories under “Mental Health” are “Enhancing Psychological Well-being”, “Anxiety and Depression”, and “Symptom Management in Mental Disorders”. Subcategories under “Exposure Therapy” are “Fear of Specific Phobias”, “Fear of Medical Procedures”, “Social phobias”, “Other phobias”. For "Mental Health and Cognitive Functioning," the subcategories are “General Public” and “Elderly Population”. “Cognitive Functioning” is divided into “Enhancing Cognitive Performance and Attention”, “Cognitive Health in Aging”, “Cognitive Performance in Clinical Populations”. The "Health" category comprises “Encouraging Physical Activity”, “Psychological Aspects of Eating and Body Behaviour”, and “Reducing Mental Symptom Discomfort”. Under "Social Well-being," the subcategories ” are “Enhancing Facial Emotion Processing” and “Enhancing Social Functioning”; the category of “Behavioural and Attitudinal Change” is not divided further at this stage; total number of subcategories (excluding the “Behavioural and Attitudinal Change” category) = 16. The numbers adjacent to each subcategory name, following the ‘=’ sign correspond to the number of studies in that subcategory.

Categories Identified

Mental Health

The “Mental Health” category consists of 79 articles divided into three subcategories: “Enhancing Psychological Well-being”, “Anxiety and Depression”, and “Symptom Management in Mental Disorders”.

The first subcategory, “Enhancing Psychological Well-being”, includes 31 studies further classified based on three elements, (a) psychological well-being construct being examined , (b) specific populations being targeted, or (c) specific situations that may impact psychological well-being. The psychological well-being constructs identified led to the following subcategories:

"Mood States and Emotional Regulation" (n=9), "Self-Perception and Agency" (n=7), "Stress and Coping" (n=5), and "Quality of Life and Functionality" (n=2). Articles within the specific populations category focused on improving psychological well-being in older adults (n=4). The remaining studies from the "Enhancing Psychological Well-being" subcategory aimed to improve mental health in individuals undergoing extreme situations or situated in extreme environments (n=4).

Stress and Coping. Symptoms of stress were primarily evaluated utilising self-report measures, at times complemented by cardiovascular measures. Among the self-report tools utilised, only two questionnaires were employed in multiple studies: the *Perceived Stress Scale (PSS),* in two alternate versions, and the *Positive and Negative Affect Schedule (PANAS).* Related psychological constructs measured alongside stress include mindfulness, psychological well-being, common mental health problems (including anxiety, depression, and somatisation), burnout, emotional regulation, and self-compassion. Additionally, the cardiovascular measures used, including heart rate (HR), heart rate variability (HRV), blood pressure (BP), and electrocardiograms (ECG) allowed for continued monitoring of the sympathetic nervous system (SNS) activity throughout the intervention, which indicated fluctuations in stress levels or emotionally stimulating experiences.

Mood States and Emotional Regulation. Emotional regulation, also referred to as the regulation of mood states, is generally examined through self-report measures. There is no identifiable trend in specific tools used across studies, as they all utilise similar but not identical questionnaires and scales. In addition to tools measuring emotional regulation and dysregulation directly, a broader array of related constructs, such as anxiety, depression, and affect, were measured. Several studies employed self-designed questionnaires to evaluate subjective mood

state changes and affective experiences post-intervention. Although less common, behavioural and physiological measures were also identified across the studies. These included spontaneous walking speed, the Stroop test, and user interactions with non-player characters (NPCs), possibly to reflect coping strategies for channelling nervous energy, assess response inhibition skills, and analyse emotion regulation in social situations, respectively. When used, various physiological measures were considered, including neurological measures (i.e., EEG), cardiovascular measures (i.e., HR, PPG), psychophysiological measures (i.e., GSR, SAM), gaze behaviour, and body temperature.

Quality of Life and Functionality. Studies focusing on quality of life (QoL) also examined measures of happiness, functional fitness, restorativeness, and subjective vitality. Similar to the categories above, self-report measures were consistently used to assess these constructs. However, there were conflicting approaches to the use of behavioural and physiological measures; one study employed both while the second used neither. Notably, physiological measures were particularly utilised to assess functional fitness.

Self-Perception and Agency. This subcategory comprises interventions aimed at altering an individual's sense of self, particularly their sense of agency, self-efficacy, self-confidence, self-criticism and self-compassion, positive self-beliefs, and overall well-being perception. The outcome measures primarily relied on various self-report tools, with no questionnaire or scale used multiple times across the studies. Behavioural assessments were rare, with only the Marino Interview Assessment Scale (MIAS) being mentioned. Additionally, only one study utilised physiological measures, specifically blood pressure and heart rate.

Psychological Well-being in Older Adults. All studies addressing general psychological well-being in the elderly were combined into this category. The dependent variables accounted for

included mental well-being, depressive symptoms, loneliness, health-related QoL, perceived social support, emotional regulation, mental health, stress, happiness, life satisfaction, anxiety, meditation experience, and sleep quality. As with other categories, self-report tools predominated among the outcome measures. Exceptions included the *Observed Emotion Rating Scale (OERS),* semi-structure interviews evaluating technology acceptance, and eye-tracking data for physiological measures.

Mental Health in Extreme Environments. Extreme environments encompassed interventions aimed at improving mental health for individuals in situations of extreme isolation, confinement, or in abnormal environments (e.g., simulated space habitat and displaced refugees in Uganda). In addition to overall mental health outcomes and psychological well-being, specific mental health constructs examined included negative mental states and stress. All studies employed an extensive amount of self-report outcome measures to assess their respective constructs. Only one study incorporated physiological measures, specifically heart rate and electrodermal activity (EDA). Semi-structured exit interviews were used in one instance to assess the VR programme and possible improvements.

The second subcategory of the "Mental Health" category, "Anxiety and Depression", regards studies primarily focused on alleviating anxiety (n=20) or depression (n=3). Articles studying anxiety reduction as their primary construct fell into the following themes: social anxiety and public speaking anxiety (n=8), anxiety in medical procedures (n=7), and co-morbid anxiety with stress or depression (n=5). One article within the "Anxiety" subcategory did not fit into any of the aforementioned themes as it studied anxiety independently. The subcategory "Depression" encompassed only three studies, as the inclusion criteria required depression to be the primary focus of the psychological intervention.

Anxiety. Studies under this subcategory mostly addressed social anxiety, public speaking anxiety, and pre-operative anxiety, and general anxiety. Anxiety was also studied in combination with stress, mood, depression, rumination, fear, distress, and self-confidence. Various combinations of outcome measures are utilised across this category, with self-report measures being the only ones consistently employed. Common self-report tools employed across multiple studies included the *Perceived Stress Scale (PSS)* and the *State-Trait Anxiety Inventory (STAI)*. For studies concerning public speaking, behavioural assessments involved delivering a speech or presentation observed by researchers. Other behavioural measures varied between third-party reports, the success of specific tasks (e.g., first-attempt intravenous insertion success), or self-monitoring and recording of pre-determined behaviours across a specific time span. Furthermore, physiological measures included cardiovascular measures, body measurements, psychophysiological measures, biochemical and hormonal measures, and neurological measures. Interestingly, three studies utilised different methods to assess cortisol levels as a measure of stress.

Depression. Depression is studied independently or in conjunction with anxiety, executive function, or distress. There was no observable trend or pattern in the outcome measures used to evaluate depression and its co-existing constructs, except for the constant employment of self-report measures. One behavioural measure was employed, the *Verbal Fluency Task (VFT)*, to evaluate executive functioning and one physiological measure assesses changes in haemoglobin.

Moreover, the final category, “Symptom Management in Mental Disorders”, comprises 25 studies and is further divided based on the mental disorder being treated by the XR psychological intervention: “Reducing Psychotic Disorder Symptoms”, “Reducing PTSD Symptoms”, “Reducing Autism Disorder Symptoms”, and “Substance Use Disorders”. It is important to note that interventions regarding the social functioning of individuals with mental disorders were not

classified under the “Mental Health” category, but in the “Social Well-being” category further described below.

Reducing Psychotic Disorder Symptoms. Psychotic disorder symptoms targeted by the included studies involved distress of auditory hallucinations, perceived voice power, acceptance of voices, ability to respond to the voices in an assertive manner, and delusion severity. Other psychological constructs studied alongside the psychotic symptoms included social and daily functioning, depression, and quality of life. Self-report measures were used almost exclusively across the studies, with repeated use of the *Psychotic Symptoms Rating Scales (PSYRATS).* Two studies employed behavioural measures, and none used physiological measures.

Reducing PTSD Symptoms. Studies focusing on PTSD symptoms generally aimed to reduce overall symptom severity without specifying individual symptoms targeted. In addition, these studies explored the symptomatology in combination with social isolation, fear, anger, depression, presence of moral injury, intrusion frequency, self-rated vividness, and emotionallity of recalled memories. Assessment tools encompassed various self-report measures and behavioural assessments. Key PTSD-specific self-report measures used repeatedly included the *PTSD Checklist for the DSM-5 (PCL-5),* the *PTSD Checklist Military Version (PCL-M),* the *Military Injury Symptom Scale– Military Short Form (MISS-M-SF),* and the *Patient Health Questionnaire (PHQ)*. Non-PTSD specific self-report measures evaluated anxiety, quality of life, physical health, executive function ratings, depression, functionality, dissociation, resilience, alcohol use, and emotion regulation. The *Clinically Administered PTSD Scale for DSM-5 (CAPS-5),* was prominently used for behavioural assessment, while daily self-monitoring of PTSD symptoms (e.g., nightmares, flashbacks, physical rage episodes, etc) and self-reports of memory vividness and emotionality in a recalling and visualisation task, were employed to a lesser extent.

One study employed physiological measures, namely electromyography, galvanic skin response, electrocardiogram and respiratory rate.

Reducing Autism Spectrum Disorder (ASD) Symptoms. This category comprises only two studies, each focusing on different aspects of ASD. One study addressed symptom severity, employing both self-report and behavioural measures to comprehensively assess ASD-related behaviours. The second study aimed to manage ASD symptoms, particularly stress levels, to enhance performance in job interviews. Unlike the first study, it did not use self-report measures but incorporated PPG and EDA data.

Substance Use Disorders. Addictive behaviours targeted in these interventions included severe alcohol use, gambling behaviour, MA-dependency, smoking dependency, and general substance use disorders. Craving was the dependent variable most frequently targeted across these studies, followed by inhibitory control, relapse rates, cognitive functioning and addiction memory. Each study employed self-report questionnaires to evaluate the extent of addiction, often complemented by behavioural assessments or physiological measures, although rarely all three simultaneously. Interventions focused on alcohol use employed various cognitive function tests and assessments on alcohol-seeking behaviours. The study on substance use disorders utilised a unique behavioural measure, urine drug testing, to monitor relapse. Similarly, physiological measures such as breathalyser and blood test were used to track substance use, alongside cardiovascular measures, eye movement measures, and respiration rates.

Most of the studies within the “Mental Health” subcategory utilise self-report questionnaires as their primary outcome measure. Behavioural measures are used to a lesser extent, followed by physiological measures. Table 1 provides specifics on the distribution of outcome measure types across all subcategories. Furthermore, Appendix F contains a detailed list of all

outcome measures pertaining to each article, organised in accordance to the subcategories mentioned above to facilitate the comparison of measures used for each dependent variable studied.

Table 1

Distribution of Studies in "Mental Health" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Enhancing Psychological Well-being	97%	23%	29%
Anxiety and Depression	96%	43%	48%
Symptom Management in Mental Disorder	88%	68%	20%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Exposure Therapy

"Exposure Therapy" encompasses all psychological interventions that utilize XR to reduce fear and anxiety through gradual exposure to the respective phobia. Among the 14 studies found in the literature involving exposure therapy, eight dealt with specific phobias, three focused on reducing fear related to medical procedures, and three addressed other phobias. The specific phobias identified included acrophobia (n=2), aviophobia (n=2), arachnophobia (n=1), vehophobia

(n=1), nyctophobia (n=1), and claustrophobia (n=1). Fears related to medical procedures regarded dentophobia (n=2), a fear of dentists, and pre-operative anxiety (n=1). The "Other Phobias" subcategory included exposure therapy interventions on reducing public speaking anxiety (n=1), a fear of contamination and disgust (n=1), and combat-related PTSD (n=1).

Fear of Specific Phobias. To examine the reduction of fear in specific phobias, self-report measures were predominantly used, followed by behavioural assessments, and lastly physiological measures. The specific phobias addressed in exposure therapy interventions included acrophobia (fear of heights), arachnophobia (fear of spiders), aviophobia (fear of flying), vehophobia (fear of driving), nyctophobia (fear of darkness), and claustrophobia (fear of enclosed spaces). The severity of these phobias was assessed using tailored questionnaires specific to each phobia, evaluating attitudes, expectations of danger, and distress levels during hypothetical approach behaviours. In addition, some studies examined related constructs such as depression, maladjustment, cognitive functioning, anxiety sensitivity and medical health. Behavioural measures involved clinical interviews for screening for diagnostic criteria of comorbid disorders. Heart rate was measured in one study.

Fear of Medical Procedures. The medical procedures included in this category focused on dental work and surgery, aiming to reduce dentophobia and pre-operative anxiety. Self-report measures were primarly used across all interventions, with one study employing all three types of outcome measures. The *Modified Dental Anxiety Scale* was the primal self-report questionnaire used for dentophobia, and various scales assessing anxiety levels and pain experience or expectations were used to evaluate pre-operative anxiety. The *Behavioural Avoidance Test (BAT)* was used in combination to heart rate measures for evaluating dentophobia.

Other phobias. This final subcategory of "Exposure Therapy" encompasses various phobias, namely public speaking anxiety, contamination-related obsessive compulsive disorder (C-OCD) and combat-related PTSD. There were no consistent self-report measures across the studies, as each aimed to assess the severity of the specific phobia studied. For public speaking anxiety, measures included self-reports of fear of negative evaluation, a behavioural assessment of anxiety severity, and physiological measures such as heart rate and electrodermal activity. Similarly, combat-related PTSD studies utilised self-report measures and behavioural assessments to evaluate symptom severity and frequency, alongside measures of depression and anxiety.

The "Exposure Therapy" category presents a varied use of outcome measures and the use of multiple outcome types per study. All studies utilise at least self-report measures to assess changes in phobia levels. Behavioural measures of approach/avoidance were also widely used across all subcategories, while physiological measures were employed half as often. Refer to Table 2 for the frequency of studies per outcome measure type and to Appendix G for the exhaustive list of measures used.

Table 2

Distribution of Studies in "Exposure therapy" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Fear of Specific Phobias	100%	38%	13%
Fear of Medical Procedures	100%	33%	33%
Other Phobias	100%	66%	66%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Mental Health and Cognitive Functioning

The "Mental Health and Cognitive Functioning" category was determined to include psychological interventions aimed at reducing mental health symptoms, such as anxiety, depression, and stress, while simultaneously improving cognitive and executive functions. The six studies in this category were further differentiated by the type of population targeted, specifically those addressing the general public (n=3), and those focusing on older adults in residential care (n=3). It is important to note that studies concerning dementia in older adults were not included in the literature review, as interventions for medical populations fall outside the scope of this review.

General Population. The mental health constructs targeted in these interventions included anxiety, depression, stress, and mood. Furthermore, cognitive functioning aspects addressed cognitive flexibility, working memory, daily living performance, cognitive performance, and

executive function. Generally, the self-report measures were used to evaluate the mental health constructs, while the behavioural assessments focused on measuring cognitive functioning. Heart rate and spontaneous eye blink rate (sEBR) were utilised as indicators of sympathetic nervous system arousal and of dopaminergic system changes, respectively.

Elderly Population. Mental health variables examined among the elderly population included general psychological well-being, feelings of apathy, loneliness, and depression. The *Apathy Evaluation Scale (AES)* was the only self-report measure used in more than one study. In addition to the mental health construts, interventions targeted cognitive functioning, cognition and verbal fluency. Unlike the subcategory concerning the general population, cognitive functioning aspects in the elderly were evaluated using self-report measures, specifically cognitive examinations. Behavioural measures assessed activity levels and included verbal tasks such as phonemic and semantic tasks. One study employed galvanic skin response and heart rate variability measures.

As shown in Table 3, all three outcome measure types in the "Mental Health and Cognitive Functioning" category are employed at least twice across the studies. Consistent with previous categories, all interventions used self-report questionnaires and complemented the self-reported data with either behavioural or physiological measures. For a comprehensive list of the specific outcome measures utilised, refer to Appendix H.

Table 3

Distribution of Studies in "Mental Health and Cognitive Functioning" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
General Population	100%	66%	66%
Elderly Population	100%	66%	33%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Cognitive Functioning

The category "Cognitive Functioning" comprises 15 articles, further classified into three subcategories. The first subcategory, "Enhancing Cognitive Performance and Attention", concerns nine studies that target different aspects of executive function in the general population.

Enhancing Cognitive Performance and Attention. Cognitive functioning constructs examined in these interventions included cognitive performance, self-perceived stress, general and selective attention, creativity, reaction time, decision-making, visual search behaviour, and inhibitory control. Conversely to the previous categories, self-report measures were infrequently used to measure cognitive improvement; the only self-report measure indicated focused on stress severity. Stroop tasks were consistently used to examine cognitive performance, selective attention and inhibitory control. Decision-making was assessed based on participant's movement decisions

during a VR simulated soccer game. All studies utilised physiological measures, specifically EEG, BP, HRV, EDA, and eye tracking.

The second subcategory, "Cognitive Health in Aging", is reserved for studies focusing on cognitive stimulation to improve cognitive functioning among the elderly, without a special emphasis on mental health (n=2).

Cognitive Health in Aging. The two studies in this category focused on executive functioning, which was primarly examined via cognitive assessments. Psychological constructs studied alongside executive functioning included levels of depression and life satisfaction. Only one study employed a behavioural measure while no physiological measures were used.

The final subcategory encompassess psychological interventions aimed at improving cognitive functioning among distinct clinical populations (n=9). Given the unique needs and challenges of functioning for each disorder, further clusters were created based on the clinical population targeted: "Schizophrenia" (n=3), "Attention-deficit/hyperactivity disorder (ADHD)" (n=2), "Autism Spectrum Disorder (ASD)" (n=2), and "Post-traumatic Stress Disorder (PTSD)" (n=2).

Attention--Deficit/Hyperactivity Disorder (ADHD). The cognitive performance aspects of ADHD evaluated in these studies included response inhibition and sustained attention vigilance. These dependent variables were predominantly examined using self-report measures; however, none of the questionnaires were consistent across studies. Both interventions employed alternative types of performance tests to study response inhibition, including a stop-signal response time (SSRT) task and another that measured omissions, commissions, response time and motor activity. No physiological measures were used in either study.

Autism Spectrum Disorder (ASD). ASD cognitive interventions focused on improving overall cognitive functioning and managing symptoms in order to facilitate job attainment. It is important to note that the job-focused intervention is applicable across various disorder diagnoses, including mood disorders, PTSD, schizophrenia, and ASD. Both interventions employed psychometric batteries, comprising both existing and self-designed tests, and incorporated behavioural assessments. The intervention aimed at improving ASD cognitive functioning employed two behavioural tests and EEG measurements. The job-focused intervention recorded various behavioural and response measures during the mock interview process but did not employ any physiological measures.

Schizophrenia. Cognitive interventions for individuals with schizophrenia aimed to enhance theory of mind (ToM), executive functioning, and interviewing self-confidence. The majority of outcome measures utilised were behavioural, followed by self-report measures, and lastly physiological measures. The only consistent behavioural assessment between the two non-interview-based interventions was the *Wisconsin Card Sorting Test (WCST),* which was used in two different versions. Similarly, two of the three interventions used the *Positive and Negative Syndrome Scale (PANSS)* to evaluate schizophrenia severity. One study utilised EEG measures. Related psychological constructs studied through self-report tools include measures of intelligence, neuropsychological status, quality of life, and self-reported knowledge and self-efficacy in performing sales-related activities. Interviewing self-confidence was assessed through neuropsychological status, emotion recognition, and a self-designed confidence scale. Behavioural measures included a job interview role play objectively scored by an observer and an emotional perspective-taking task.

Post-Traumatic Stress Disorder (PTSD). Cognitive performance interventions focused on PTSD populations utilised self-report measures and behavioural measures comparably, while physiological measures were not as prevalent. The specific dependent variables targeted, apart from cognitive functioning, included psychological symptoms of PTSD, alcohol use, alcohol craving and subjective distress. The emphasis on PTSD symptomatology necessitated self-report measures on symptom experience in addition to cognitive examinations. Heart rate variability was used to compliment these outcome measures. Moreover, a focus on the effect of alcohol dependency in cognitive performance of individuals with PTSD involved exploring alcohol use habits, and potential underlying mental health issues such as like depression and anxiety. Numerous behavioural measures were employed to evaluate cognitive functioning.

As reflected in the retrieved literature on cognitive functioning interventions, behavioural measures are the primary outcome measures used to assess cognitive improvement, closely followed by self-report measures. While physiological measures are less common, they are used to a larger extent across the first subcategory, “Enhancing Cognitive Performance and Attention”. Table 4 indicates the specific frequencies of outcome measures used across each subcategory while Appendix I provides details on the specific questionnaires, tests, and measures utilised per outcome type and intervention.

Table 4

Distribution of Studies in "Cognitive Functioning" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Enhancing Cognitive Performance and Attention	25%	75%	100%
Cognitive Health in Aging	100%	50%	0%
Cognitive Performance in Clinical Population	100%	100%	33%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Health

Studies classified under “Health” encompass psychological interventions related to encouraging physical activity among various age groups, eating behaviour and body image, and alleviating medical symptom discomfort. The subcategory “Encouraging Physical Activity” includes psychological interventions focused on two themes: exercise adherence (n=4) and perceived exertion and motivation (n=3).

Physical Activity and Exercise Adherence. In general, interventions aimed at encouraging exercise behaviour focused on self-determination, adherence and motivation. One study, however, had an emphasis on changes in multi-sensory processing as a result from exercise. Predominantly, self-designed self-report measures were used in this category to assess participants’ intrinsic motivation, perceived social support to exercise and their experiences with

at-home physical therapy. These measures included assessments of pain perception, attention demand, learning, and control perception, in addition to intervention-related measures. Behavioural measures included exercise performance and activity levels, recorded by wearables, in addition to a behavioural assessment task. Physiological measures included real time motion tracking through the wearables.

Perceived Exertion and Motivation. Measures of perceived exertion were coupled with assessments of self-efficacy, enjoyment, intrinsic motivation, subjective vitality, and exercise intentions. All types of outcome measures were utilised across these interventions, with the majority classifying as self-designed self-report measures assessing the aforementioned psychological constructs. Behavioural measures included frequency of exercise, mean power output during exercise, and eating behaviour habits. Furthermore, various physiological measures were employed, specifically cardiovascular metrics (blood pressure and VO2 level), and body measurements (weight, height, and BMI).

The second subcategory, "Psychological Aspects of Eating and Body Image," covers studies examining body image attitudes and behaviours (n=8) and eating behaviour and perception (n=2). The final subcategory within the "Health" category includes two articles focused on reducing distress and discomfort, one in sleep-deprived individuals and one in patients with tinnitus.

Body Image Attitudes and Behaviours. Studies focused on body image encompassed dependent variables such as body image disturbance (BID), body image attitudes, estimation of body size and body parts, eating disorder psychopathology, and the perceptual boundary of body classifications (fat versus thin). Despite the common focus and vast use of self-report measures, only one scale, *the Eating Disorders Examination Questionnaire (EDE-Q)*, was used in multiple

studies. Other self-report measures assessed related-constructs such as weight concerns, body satisfaction/dissatisfaction, body image avoidance, depression severity, self-esteem, and scales specific to eating disorders. Behavioural assessments encompassed approach and avoidance tasks, clinical interviews to evaluate the presence an eating disorder, self-perception estimations of body part sizes, and passive and active weight estimation tasks. Notably, measures of embodiment were highly valuable across these interventions. The sole physiological measure used was body mass index (BMI), which was recorded in several studies.

Eating Behaviour and Food Perception. These interventions aimed to study responses to food cues (FCR) and self-efficacy in changing eating behaviour based on dietary carbon footprint. Self-report measures were primarily used, followed by physiological measures, with no behavioural measures employed. Craving was assessed through self-report and complemented by the physiological measure of salivation. Additionally, food frequency, self-efficacy, intentions, and psychological distance were evaluated through self-report measures.

Reducing Medical Symptom Discomfort. Interventions aimed at relieving psychological symptoms of sleep deprivation and tinnitus predominantly used self-report measures. Both interventions employed the *Pittsburg Sleep Quality Index (PSQI)* to assess sleep quality. Additionally, the tinnitus-related intervention incorporated assessments examining quality of life, anxiety, depression, and mood states. This intervention also utilised EEG measurements for data collection.

Health-related dependent variables are most commonly studied using self-report measures, according to the data retrieved from the literature. While used to a lesser extent, behavioural and physiological measures are comparably utilised alongside the self-report questionnaires. Refer to

Table 5 for details on the number of studies employing each outcome measure. Furthemore, a comprehensive list of all outcome measures is provided in Appendix J.

Table 5

Distribution of Studies in "Health" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Encouraging Physical Activity	100%	71%	57%
Psychological Aspects of Eating and Body Image	100%	63%	25%
Reducing Medical Symptom Discomfort	100%	0%	50%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Social Well-being

The “Social Well-being” category consists of 25 studies organised into two subcategories, namely “Enhancing Facial Emotion Processing” and “Enhancing Social Functioning”. The first subcategory encompassed psychological interventions focused on improving emotion recognition and responding to emotion expression (n=5). Four out of five of the studies in this subcategory involved clinical populations, specifically individuals with autism (n=3) and schizophrenia (n=1).

Impact of Virtual Abuse on Emotional Processing. No patterns or trends could be identified given that only one study is classified under this subcategory.

Emotion Recognition in Clinical Populations. Autism was the primary clinical population studied in these interventions, with the exception of one study that focused on schizophrenia. All of these interventions aimed to assist in emotion processing and response, as well as theory of mind. Consistent with previous categories, several self-report measures were employed in all interventions but none were used across multiple studies, except for one: the *Social Communication Questionnaire (SCQ).* Behavioural measures were also frequently used, emphasising tasks of emotion identification and social attribution. Comparably, physiological measures appeared to be of high value, spanning eye tracking measures (gaze patterns, blink rate, pupil diameter), cardiovascular measures (ECG, PPG, BOLD responses), body measurements (skin temperature) and psychophysiological measures (GSR).

The second subcategory is comprised of four themes: improving social connectedness (n=6), improving social cognition and functioning in individuals with psychotic disorders (n=6), enhancing social skills in individuals with Autism Spectrum Disorder (ASD) (n=5), and increasing empathy and perspective-taking (n=3).

Social Connectedness. Studies concerning social connectedness evaluated joint-decision making, social inclusion, self-perceived social closeness, social cognition and functioning, mimicry, and prosocial bystander behaviour. The majority of outcome measures were self-report assessments, addressing the aforementioned dependent variables as well as attitudes towards objects, mental well-being, affect, depression, anxiety, stress, and self-esteem. Behavioural measures included physical proximity in VR to others, instances of mimicry, tasks assessing awareness of social inferences, and self-reporting of social interactions. The bystander intervention

deviated from these patterns, focusing on self-efficacy as a bystander, alcohol use behaviours, and attitudes toward sex, rape, and empathy. No physiological measures were used.

Empathy and Perspective-taking. All interventions in this category exclusively relied on self-report measures. One study focused on empathy utilized validated scales such as the *Affective Empathy Scale*, while others employed self-designed questionnaires directly related to the context of the VR intervention.

Social skills in Autism Spectrum Disorder (ASD). The social skills specific to ASD targeted in these interventions included social communication, emotional expression, non-verbal communication, and social interaction with eye contact. All outcome measures were used at relatively similar frequencies. Self-report measures were predominantly of validated scales, with one instance of a self-designed questionnaire. Two self-report tools were used across the studies, namely the *Social Responsiveness Scale (SRS)*, and the *Social Communication Questionnaire (SCQ)*.

Furthermore, behavioural measures identified primarily focused on diagnostic assessments for ASD or tests to assess ASD symptoms and behaviours. The most frequently used physiological measures across the studies were eye movement measures, including eye gaze monitoring, blink rate, mean blink rate, pupil diameter, sum of fixation counts, and average fixation duration. Additionally, cardiovascular measures (ECG and PPG), body measurements (temperature), psychophysiological measurements (GSR and RR), and neurological measurements (EEG and EMG) were also employed.

Social Cognition and Functioning in Psychotic Disorders. The social cognition interventions for psychotic disorders lacked specificity regarding the skills being targeting. These vague dependent variables addressed included social skills, social cognitive abilities, social

functioning, social attention, social participation and quantity and quality of social contacts. However, specific constructs underlying the dependent variables can be inferred from the self-report measures utilised across the studies: social avoidance and distress, emotional intelligence, empahy, global functioning, quality of life, perception of hostility in ambiguous scenarios, self-esteem, internalised stigma of mental illness and dating assertiveness. Additionally, all studies employed tools to identify the severity of psychotic symptoms. Behavioural assessments aimed to measure social cognition through emotional recognition tasks, face memory tasks, verbal fluency tests, and daily self-reports on social interactions at random moments. Physiological measures were not employed.

Regarding outcome measures, both subcategories within the "Social Well-being" category show a preference for self-report measures. Given the difference in number of studies across the categories, the frequencies are not directly comparable. The first subcategory, "Enhancing Facial Emotion Processing", utilised each outcome measure type to a similar extent, in the following order: self-report measures, physiological measures, and finally, behavioural measures. In contrast, the second subcategory, "Encouraging Social Functioning", exhibited a disproportionate preference for behavioural measures compared to physiological measures. For exact values for each outcome measure type, refer to Table 6. A detailed list of the specific measures used per outcome type is provided in Appendix K.

Table 6

Distribution of Studies in "Social Well-being" Category by Outcome Measure Type

Subcategory	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Enhancing Facial Emotion Processing	100%	60%	80%
Enhancing Social Functioning	95%	5%	15%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Behavioural and Attitudinal Change

The final main category, “Behavioural and Attitudinal Change,” encompasses three distinct themes: behavioural change, attitudinal change, and risk behaviour and safety. Studies under the behavioural change theme primarily address maladaptive social behaviours, specifically violence, aggression, and bullying incidents (n=3). There is one exception in which the study investigates behaviour change regarding hot water use.

Behavioural change interventions targeted encouraging bystander behaviour and energy-saving behaviours, as well as preventing bullying and aggression. Self-report measures were used at a comparable rate to behavioural measures, while physiological measures were absent. The majority of self-report measures were validated scales specific to each study, with no consistent use of any outcome measure across studies. encompassing observed body language in social interactions within VR scenarios, third-party reports of behaviour, and measurements such as

water temperature and quantity used. There was no use of physiological measures across this subcategory.

The second theme, attitudinal change, consists of two psychological interventions: one aimed at improving beliefs of individuals with negative attitudes towards vaccination and another concerning organisational training.

Interventions focused on attitudinal change addressed perception, belief formation, and organisational training. Both studies exclusively utilized self-report measures, comprising self-designed questionnaires directly related to the attitudes addressed, assessments of knowledge acquired through the intervention, and intentions to change based on the acquired knowledge.

The last theme concerning risk behaviour and safety includes various contexts (n=5). Two of the interventions in this category involve reducing risk behaviourn specifically related to driving. The remaining three articles focus on increasing safety awareness in distinct situations, including injury prevention in occupational environments, street crossing safety, and fire safety in medical environments.

Interventions in this subcategory targeted risk driving behaviour and safety skills in occupational environments and street crossing. Those focusing on reducing risk behaviours while driving relied exclusively on behavioural measures, with the exception of one study that used the *Driver Behaviour Questionnaire (DBQ)* and the *Emotional Arousal Scale*. The behavioural measures assessed real-life DUI behaviours and risk-taking tendencies in virtual reality (VR) environments. Studies aiming to increase safety awareness utilised both self-report and behavioural measures. Self-report measures were researcher-designed to assess specific knowledge of safety guidelines, while behavioural measures evaluated the application of these guidelines in VR-simulated emergency scenarios.

Self-report measures and behavioural measures were used almost as frequently in behavioural/attitudinal change interventions. Conversely, the utilisation of physiological measures was nonexistent. Precise frequencies per self-report measures and behavioural measures are presented in Table 7. Additionally, specific assessments used per outcome measure type are outlined in Appendix L.

Table 7

Distribution of Studies in "Behavioural and Attitudinal Change" Category by Outcome Measure Type

	Outcome Measure Type		
	% Self-report	% Behavioural	% Physiological
Behavioural and Attitudinal Change	73%	64%	0%

Note. Each percentage stated above is directly derived from the total number of studies in the respective subcategory. Studies that utilize more than one outcome measure type are included in the count for each outcome type. Percentages were rounded to the nearest tenth.

Ultimately, a summary list of the seven main categories, along with their corresponding subcategories and further classifications, can be found in Appendix M for simplicity.

Outcome Measures Identified

A total of 491 outcome measures were identified across the 167 studies on XR-based psychological interventions and were categorised into two main groups: “Intervention-related

measures" and "Participant-related measures". In this literature review, intervention-related measures encompass assessments or evaluations focused on the VR experience itself, independent of the psychological construct or dependent variable being studied. Constructs included under this category are measures of perceived presence, immersion, satisfaction, cyber-sickness symptoms, and usability. In total, 19 intervention-related measures were identified across all studies in the literature review (Appendix N). When listed as outcome measures in the respective result tables across the literature review categories found in the Appendices F through L, intervention-related measures are indicated by the use of an asterisk (*).

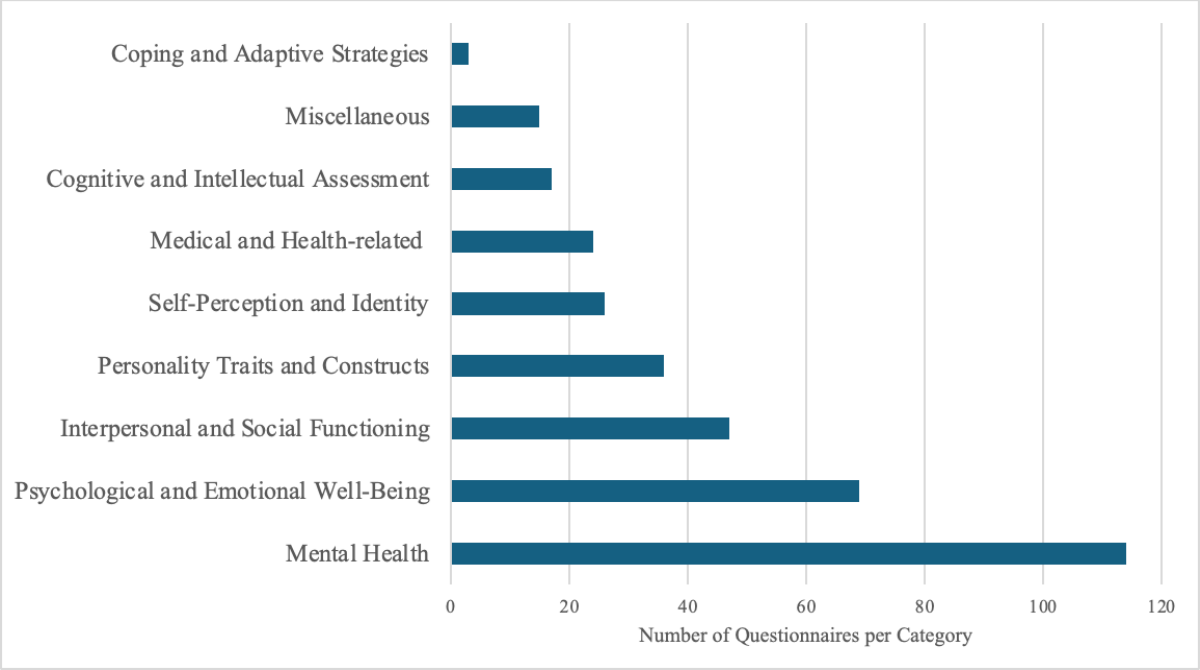
The second category, "Participant-related measures", comprised three subcategories directly corresponding to the types of outcome measures: self-report, behavioural and physiological. Self-report measures were the most extensively utilised across all XR-based psychological interventions examined, as reflected in the 344 distinct assessments and tests classified under this category. The 344 self-report measures identified were further classified into nine categories: "Mental Health" (n=114), "Psychological and Emotional Well-Being" (n=69), "Interpersonal and Social Functioning" (n=47), "Personality Traits and Constructs" (n=36), "Self-Perception and Identity" (n=26), "Medical and Health-related" (n=24), "Cognitive and Intellectual Assessment" (n=17), "Miscellaneous" (n=15), and "Coping and Adaptive Strategies" (n=3). Refer to Figure 6 for a visual representation of the frequencies. Within each category, assessments measuring the same psychological construct were combined into subgroups, find a detailed list of these subcategories in Table 8.

It should be noted that two self-report measures involved recording the occurrence of "adverse events" during the intervention period. Due to the extensive list of adverse events, these are not specified within the main tables. Refer to Appendix O for the detailed list of adverse effects

respective to each study. Additionally, Appendix P provides several definitions relevant to the outcome measures identified.

Figure 6

Distribution of Self-Report Measures Across Categories



Note. The figure above depicts the self-report measure subcategories in descending order according to the number of measures within each category.

Table 8

Classification of Self-Report Measures into Categories and Subcategories and the Total Number of Questionnaires Identified for Each Psychological Construct

No.	Category	Construct measured	Number of Questionnaires	Percentage (%)
1	Mental health	Addiction-related	13	4%
		Attention Deficit Hyperactivity Disorder (ADHD)-related	3	0.9%
		Autism Spectrum Disorder (ASD)-related	10	3%
		Anxiety	22	6%
		Borderline Personality Disorder (BPD)-related	1	0.3%
		Depression	13	4%
		Mental Illness Assessments	3	0.9%
		Obsessive Compulsive Disorder (OCD)-related	1	0.3%
		Phobia-related	17	5%
		Psychotic-related	20	6%
		PTSD	6	2%
		Trauma-related	5	1%
		Total	114	33%
2	Psychological and Emotional Well-Being	Emotional functioning	18	5%
		Fear	4	1%

		Happiness and Vitality	3	0.9%
		Overall Well-Being and Functioning	22	6%
		Quality of Life	17	5%
		Stress	3	0.9%
		Worry	2	0.6%
		Total	69	20%
3	Interpersonal and Social Functioning	Belonging	3	0.9%
		Bullying-related	2	0.6%
		Bystander Behaviour	7	2%
		Family Functioning	1	0.3%
		Mood States and Affect	6	2%
		Public Speaking/ Communication	5	1%
		Social Functioning	23	7%
		Total	47	14%
4	Cognitive and Intellectual Assessment	Cognitive Examinations	9	3%
		Intelligence	8	2%
		Total	17	5%
5	Self-Perception and Identity	Self-perception	7	2%
		Body-image, Body Perception, and Food Perception	19	6%

		Total	26	8%
6	Coping and Adaptive Strategies	Coping Strategies	1	0.3%
		Meditation and Mindfulness	2	0.6%
		Total	3	0.9%
7	Medical and Health-related	Medical	8	2%
		Pain	5	1%
		Sexual Risk	5	1%
		Exercise-related	5	1%
		Sleep	1	0.3%
		Total	24	7%
8	Personality Traits and Constructs	Agency	1	0.3%
		Aggression and Anger	8	2%
		Assertiveness and Self-Confidence	4	1%
		Attitudes and Beliefs	4	1%
		Ego	1	0.3%
		Engagement with Beauty	1	0.3%
		Helping Intentions	1	0.3%
		Impulsiveness	2	0.6%
		Locus of Control	1	0.3%
		Motivation	1	0.3%

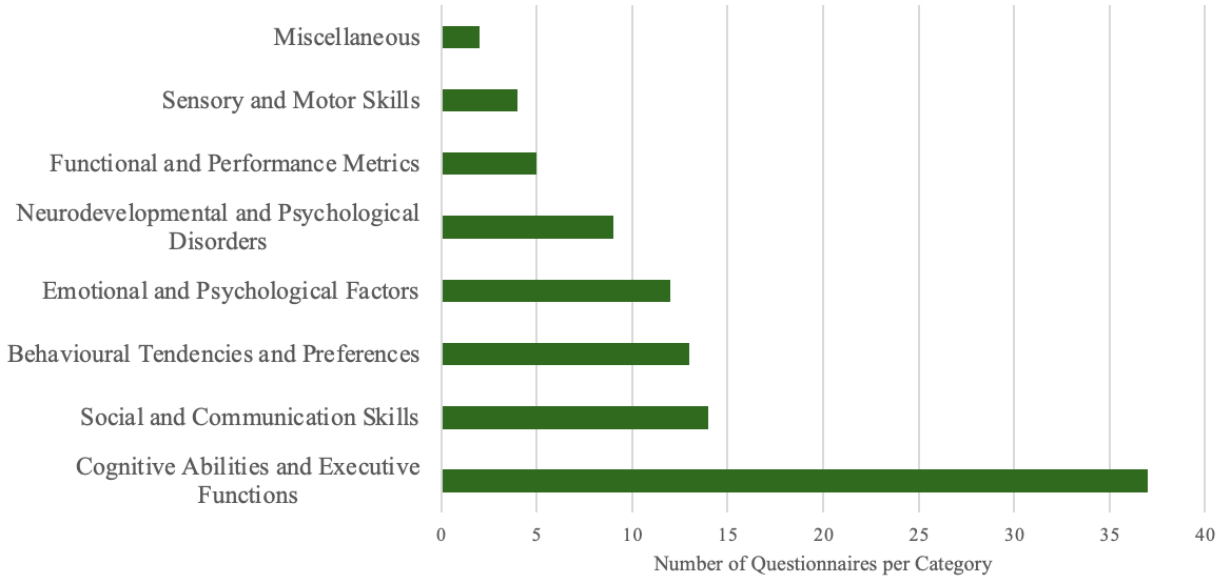
		Optimism	1	0.3%
		Personality	2	0.6%
		Perspective-taking	1	0.3%
		Resilience	2	0.6%
		Sympathy and Empathy	6	2%
		Total	36	10%
9	Miscellaneous	Academic Engagement	1	0.3%
		Insights into Experience	3	0.9%
		Self-Reported Behaviour	6	2%
		Stigma of Mental Illness/ Mental Health Services	3	0.3%
		Types of Measurement	2	0.6%
		Total	15	4%

Note. The table above indicates the total number of questionnaires assessing each psychological construct. Questionnaires that are used and validated for multiple psychological constructs (e.g., Hospital Anxiety and Depression Scale (HADS)) were included in the counts for each of the psychological constructs (e.g., "Anxiety" and "Depression"). This approach was taken only when the questionnaire was specifically designed to assess each of the included constructs independently. For this reason, the sum of all counts in the table is 351, rather than the initially identified 344 self-report measures. Percentages were rounded to the nearest tenth, except for percetages smaller than 0, which were rounded to the nearest hundredth.

The “Behavioural Measures” category followed in preference, with a total of 94 unique tasks or observational evaluations utilised across the studies. Similarly, these assessment tools were further divided into eight categories based on the construct studied by each task. The main categories for the behavioural measures were “Cognitive Abilities and Executive Functions” (n=37), “Social and Communication Skills” (n=14), “Behavioural Tendencies and Preferences” (n=13), “Emotional and Psychological Factors” (n=12), “Neurodevelopmental and Psychological Disorders” (n=9), “Functional and Performance Metrics” (n=5), “Sensory and Motor Skills” (n=4), and “Miscellaneous” (n=2). Figure 7 illustrates the distribution of behavioural measures across the aforementioned categories. The respective subcategories and specific counts are specified in Table 9.

Figure 7

Distribution of Behavioural Measures Across Categories



Note. The figure above depicts the behavioural measure subcategories in descending order according to the number of measures within each category.

Table 9

Classification of Behavioural Measures into Categories and Subcategories

No.	Category	Construct measured	Number of Questionnaires	Percentages (%)
1	Cognitive Abilities and Executive Functions	Cognitive Profiling Tools	5	5%
		Memory Assessment Tools	9	9%
		Executive Functioning Assessment Tools	4	4%
		Processing Speed Assessment Tools	7	7%
		Cognitive Inhibition Assessment Tools	2	2%
		Response Inhibition Assessment Tools	2	2%
		Cognitive Flexibility Assessment Tools	3	3%
		Intelligence Assessment Tools	2	2%
		Verbal Working Memory and Fluency Assessment Tools	2	2%
		Creativity Assessment Tools	1	1%
		Total	37	37%
2	Emotional and Psychological Factors	Facial Emotion Recognition Assessment Tools	6	6%
		Affective Observation Tools	2	2%
		Anxiety Assessment Tools	2	2%

		Post-Traumatic Stress Disorder (PTSD) Assessment Tools	2	2%
		Total	12	12%
3	Social and Communication Skills	Theory of Mind (ToM) Assessment Tools	6	6%
		Social Cognition and Communication Assessment Tools	3	3%
		Language Assessment Tools	2	2%
		Observational Assessments by Third-Parties	3	3%
		Total	14	14%
4	Neurodevelopmental and Psychological Disorders	Autism Spectrum Disorder (ASD) Assessment Tools	2	2%
		Clinical Interviews for Psychiatric Disorder Diagnosis	3	3%
		Alcohol-Related Behaviour Assessment Tools	4	4%
		Total	9	9%
5	Sensory and Motor Skills	Visual Processing Assessments	3	3%
		Visual-Motor Integration and Preferences	1	1%
		Total	4	4%
6	Behavioural Tendencies and Preferences	Approach and Avoidance Behaviour Assessment Tools	4	4%

		Attention Assessment Tools	7	7%
		Risk-taking Assessment Tools	2	2%
		Total	13	13%
7	Functional and Performance Metrics	Functional Capacity and Skills Assessment Tools	3	3%
		Performance Assessment Tools	2	2%
		Total	5	5%
8	Miscellaneous		2	2%
		Total	2	2%

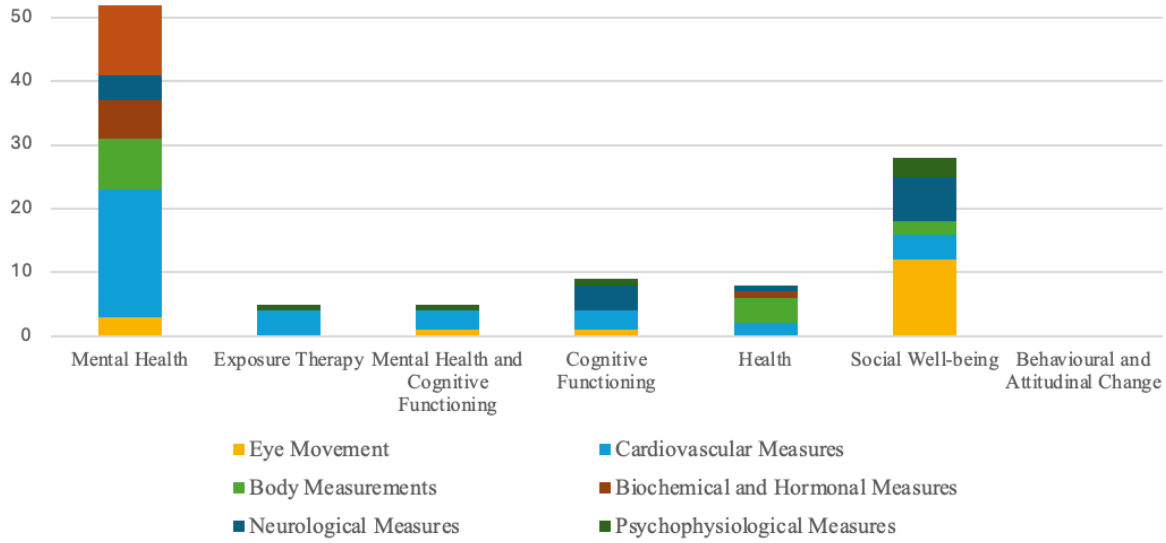
Note. The table above presents the overall count of assessment tools classified under each subcategory. Similarly to the self-report measures table, some measures are included in more than one subcategory, resulting in a total sum of 96 instead of 94. Percentages were rounded to the nearest tenth, except for percetages smaller than 0, which were rounded to the nearest hundredth.

Finally, 39 different physiological measures were employed across the interventions. These interventions were further classified into six broad categories depending on their nature: “Cardiovascular Measures” (n=36), “Eye Movement” (n=19), “Psychophysiological Measures” (n=17), “Neurological Measures” (n=16), “Body Measurements” (n=14), and “Biochemical and Hormonal Measures” (n=7). Refer to Figure 8 for a visual representation of the distributions of physiological measure categories across the main categories, based on the dependent variable.

Table 10 presents the specific physiological measures under each broad category and indicates the number of times each measure was utilised across each of the main categories.

Figure 8

Distribution of Physiological Measures Across Main Categories



Note. The stacked bar graph above illustrates the number of physiological measures utilized within each dependent variable-based category, indicating the specific types and frequencies of each measure, differentiated by colours.

Table 10

Classification of Physiological Measures into Categories and Frequencies

No.	Category	Construct measured	Frequencies per Category						
			1	2	3	4	5	6	7
1	Eye Movement	Average fixation duration						1	
		Binocular eye movement	1						
		Blink rate (BR)						3	
		Eye tracking data	1			2			
		Gaze behaviour	1					5	
		Pupil diameter (PD)						3	
		Spontaneous eye blink rate (sEBR)			1				
		Sum of Fixation Counts (SFC)						1	
		Total	3	0	1	2	0	13	0
2	Cardiovascular Measures	Blood pressure	4			1	1		
		Heart rate (HR)	9	4	2				
		Heart rate variability (HVR)	2		1	2			
		Electrocardiogram (ECG)	2					2	

		Photophethysmographic (PPG) data	1					1	
		Pulse plethysmogram (PPG)	2					1	
		VO_2 level					1		
		Total	20	4	3	3	2	4	0
3	Body Measurements	Body mass Index (BMI)	1				2		
		Height					1		
		Waist circumference (WC)	1						
		Waist-to-hip ratio	1						
		Weight	1				1		
		Body/skin temperature (SKT)	4					2	
		Total	8	0	0	0	4	2	0
4	Biochemical and Hormonal Measures	Carbohydrate deficient transferrin (CDT) levels via blood test	1						
		Salivary Cortisol levels	1						
		Gamma-glutamyl transpeptidase (GGT) levels via blood test	1						
		Haemoglobin changes	1						
		Salivary Alpha Amylase (sAA) Enzyme quantity	1						
		Cortisol levels with ELISA method	1						

		Salivation quantity					1		
		Total	6	0	0	0	1	0	0
5	Neurological Measures	Electroencephalogram (EEG)	2			4	1	1	
		Electromyography (EMG)	2					1	
		Default Mode Network Changes (DMN)						1	
		Functional Magnetic Resonance Imaging (fMRI)						1	
		Anatomical Images						1	
		Separate functional localizer						1	
		BOLD (blood oxygen level dependent) responses with fMRI						1	
		Total	4	0	0	4	1	7	0
6	Psychophysiological Measures	Electrodermal activity (EDA)	5	1		1			
		Galvanic Skin Reaction/Response (GSR)	2		1			2	
		Respiratory rate (RR)/ Respiration (RSP)	3					1	
		Self-assessment model (SAM)	1						
		Total	11	1	1	1	0	3	0
		Sum of Total Measures Used Across Each DV Category	52	5	5	10	8	29	0

Note. The columns labeled "Frequencies per Category" are numbered from 1 to 7. Each number corresponds to a specific main category referenced throughout this literature review: "Mental Health" is represented by 1, "Exposure Therapy" by 2, "Mental Health and Cognitive Functioning" by 3, "Cognitive Functioning" by 4, "Health" by 5, "Social Well-being" by 6, and "Behavioural and Attitudinal Change" by 7. The frequencies listed in the table indicate the number of times a particular physiological measure was used within each category. The totals do not match the number of studies per category since some studies did not employ any physiological measures while others utilised several.

For a comprehensive list of all 491 outcome measures, organised alphabetically, by outcome type, and by the identified categories above, with adjuncted descriptions on what construct is being measured, refer to Appendix P. Specifically, self-report measures are indicated in Appendix Q1, behavioural measures in Appendic Q2, and physiological measures in Q3.

Checklist

To enable the application of findings from the systematic review, a prototype version of a checklist tool was developed to guide future XR psychological interventions. This tool utilises a compilation of mind maps to visually represent a systematic thinking strategy. Specifically, these mind maps enable a thorough assessment of various considerations related to the dependent variable being studied, structured in an organised manner. This approach aims to help researchers gain a clear overview of the complex elements in their XR psychological interventions.

The checklist tool developed comprises eight mind maps: one screening mind map that directs the user to the appropriate section of the checklist based on their answers, and seven others corresponding to the main categories derived from the dependent variable (Appendix R). Despite the difference in content, all mind maps utilise the same elements: closed questions with their respective answer options, black lines, black arrows, blue arrows, and insight text followed by bullet points.

The mind maps include two types of closed questions: dichotomous and multiple-choice. Dichotomous questions offer the respondents of 'Yes' and 'No' response options and are followed by further questions or by insights based on the literature review's findings. Multiple-choice questions provide a range of responses, often including a 'None of the above' option followed by the possibility to specify the missing option that best fits their intervention. These questions primarily aim to gather specific information about the targeted population and refine the intervention's scope. Like the dichotomous questions, the multiple-choice questions are accompanied by additional questions or insight. However, some exceptions are observed as topics not addressed in the literature were also omitted from the checklist due to its direct reliance on the literature reviewed.

Connections between the mind maps are represented using black lines and two different coloured arrows. Solid black lines guide the user through the map, connecting the respondent's answer to a subsequent question based on the selected response. Black arrows signify questions that follow regardless of the respondent's previous answer. Blue arrows indicate insights from the literature that researchers should consider for the respective dependent variable. These insights form the concluding section of each mind map branch. Furthermore, they are aimed to be informative and not compulsory, offering guidance on how specific constructs are evaluated

or which additional constructs are commonly studied alongside a given variable. Insight text comprises a brief introductory statement followed by the insight in the form of bullet points.

The eight mind maps can be found in Appendix R. The screening mind map is added under Appendix R1, followed by “Mental Health” in Appendix R2 and “Exposure Therapy” in Appendix R3. Mind maps for the categories “Mental Health and Cognitive Functioning” and “Cognitive Functioning” are found in Appendices R4 and R5, respectively. Furthermore, Appendices R6 and R7 regard “Health” and “Social Well-being”. Lastly, the mind map for the category of “Behavioural and Attitudinal Change” can be found in Appendix R8.

Future steps regarding the improvement, implementation, and dissemination of the current prototype are discussed in the following sections of this report.

Discussion

This literature review aimed to evaluate the current landscape of XR-based interventions, specifically concerning the dependent variables being studied and the respective methods of measurement. The review revealed seven broad themes encompassing various psychological constructs consistently studied throughout XR interventions. Moreover, 491 unique outcome measures were identified, highlighting the extensiveness of the measurement approaches employed. The discussion below delves into the patterns observed regarding the use of these outcome measures used across each the dependent variables. Additionally, the significant need for a checklist to aid researchers in designing future XR psychological intervention and its implications on standardisation and comparability are recognised.

General Interpretation of Results

The comprehensive evaluation of XR-based psychological interventions revealed seven overarching themes currently studied in the field, alongside the use of distinct outcome measures. These themes encompass interventions targeting mental health and psychological well-being, cognitive functioning, health-related quality of life, eating behaviours, social functioning, and behavioural and attitudinal change across diverse populations including the general population, elderly, and those with mental disorders. Across these studies, self-report, behavioural, and physiological measures were consistently utilised to track improvements in psychological constructs. The application of XR technology facilitates the assessment of behavioural and physiological measures by providing realistic simulations within controlled environments. This capability enhances ecological validity while also encouraging the comprehensive evaluation of constructs, encompassing both subjective experiences and objective

data. Furthermore, the diversity of outcome measures identified across the literature underscores the abundance of multifaceted approaches to monitoring psychological change and improvement.

Evaluation of Dependent Variables Studied Using XR Technology

The systematic review revealed a strong preference for utilising XR technology in interventions targeting mental health. This finding can be attributed to the high global burden of mental health disorders, affecting approximately 1 in every 8 individuals worldwide according to estimates from the World Health Organisation (WHO) (World Health Organization, 2022). Mental health conditions not only have a high prevalence but also encompass broad spectrum of diagnoses, specifically when there is clinically significant disturbance in cognition, emotional regulation, or behaviour (World Health Organization, 2022). The extensive application of XR technology on mental health interventions enables greater accessibility and expands the treatment capacity, as it reduces reliance on direct psychologist involvement. This preference for using XR technology to improve mental health is not only justified by the prevalence of mental health challenges, but also provides a solution to the issue of inadequate treatment provision, which significantly contributes to years with lived disability (WHO, 2024).

It is noteworthy that the “Mental Health” category, which encompasses the highest number of XR psychological intervention, extensively overlaps with the other categories as well, emphasising mental health as the primary focus of XR interventions. This overlap is particularly meaningful for the categories “Exposure Therapy”, which targets anxiety arising from specific triggers, and “Mental Health and Cognitive Functioning”, which comprises studies that equally target mental health and cognitive constructs.

Despite the recognised potential for XR to enhance exposure therapy treatments, as evidenced by the established name VRET (Virtual Reality Exposure Therapy) for these

interventions, there were relatively few XR interventions involving exposure therapy within the extracted literature. Research comparing VRET to in vivo exposure indicates higher efficiency due to its ability to deliver treatment from any location, promote outcomes through enhanced presence from the start, and enable real-time data analysis for personalised treatment adjustments (XRHealth, n.d.). Additionally, VRET provides a more flexible delivery mechanism for exposure therapy, capable of addressing implementation barriers found in in vivo exposure while retaining effectiveness (Levy et al., 2023).

Consistent with the XR interventions identified in this review, research has demonstrated promising results from VRET in treating phobias, social anxiety disorder, obsessive-compulsive disorder, and panic disorder (XRHealth, n.d.). Research also indicates participant preference for VRET over in vivo exposure therapy. Specifically, it was found that participant concerns about undergoing exposure therapy treatment vary depending on whether they proceed with in vivo exposure or VRET. In vivo exposures often evoke concerns about exacerbating mental health issues, including increased anxiety, feelings of embarrassment or shame, and worsening of the current condition (Levy et al., 2023). Conversely, concerns related to VRET treatment comprised more of external factors, such as risks of side effects, efficacy uncertainty, and healthcare insurance coverage (Levy et al., 2023). This parallel in participant concerns highlights greater acceptance and trust in using XR technology compared to in vivo treatments. Despite these findings, this user preference for XR-based exposure therapy is not yet fully supported by real-world intervention studies.

Moreover, current literature highlights a trend towards applying XR technology in interventions focused on social functioning. Primarily, these interventions aim to enhance social connectedness in the general population and improve social cognition in clinical populations,

particularly among individuals with autism spectrum disorder (ASD) and psychotic disorders. The interest in utilising XR in interventions targeting social connectedness could be attributed to its unique capacity to facilitate naturalistic interactions in intentional environments with engineered avatars and controlled avatar behaviours. This ability provides researchers a novel method to observe social behaviours in controlled settings, which enables replicability and customisation for both longitudinal and cross-sectional study approaches.

Consistent with the results of this review, research on the use of XR for enhancing social functioning seems to be primarily concerned on its potential benefits for the autistic community. According to a systematic review by Chen et al.'s (2022), positive improvements in social interaction, acceptance, engagement, communication, emotion recognition, daily functioning, reduction in problem behaviours, attention, contextual processing, and anxiety symptoms have been observed in individuals with ASD participating in XR psychological interventions. Conversely, there is comparatively less research on the use of XR as a tool to improve social skills in individuals with psychotic disorders. Despite the lower prevalence, studies indicate that VR-based social cognition and interaction training (VR-SCIT) is promising for improving emotion perception and metacognition in individuals with schizophrenia or other psychotic disorders, in addition to increasing treatment compliance.

Evaluation of Outcome Measures Usage

Outcome measures were initially divided into two general categories: those aimed at gathering insights on the XR intervention experience and those assessing psychological change. Using intervention-related measures in combination with participant-related measures is essential, especially as XR technologies continue to emerge in the field of psychology. These intervention-related measures focus on the effectiveness of using XR as a platform for simulating

real-life scenarios to generate genuine reactions from the participants. Additionally, they are concerned with the user's subjective experience with the XR technology. The questionnaires under this category assess feelings of satisfaction, side effects experienced, acceptance, and overall ease of use. Measures of perceived presence and immersion, also assessed through intervention-related measures, provide crucial information that may strengthen or weaken the changes identified through the participant-related measures. For example, in an exposure therapy XR intervention, high feelings of presence and immersion are fundamental to evoke genuine anxiety and fear in the participants to treat it effectively. If adequate feelings of immersion are not attained, the use of XR could potentially hinder the patient's psychological treatment compared to traditional real-life exposure therapy. Thus, understanding the user's perceived immersion is essential for optimising current interventions and developing new ones.

Regarding participant-related measures, self-report questionnaires were predominant across most of the identified XR psychological interventions categories. This reliance on subjective data underscores its effectiveness for collecting data of internal psychological processes like thoughts, feelings, and behaviours. The widespread adoption of self-report measures can also be attributed to its historical precedence, given the limitations for alternative objective measures prior to recent technological advances. While subjective experiences and self-perceptions are crucial for understanding individual behaviour, complementing these with objective measures could enhance overall comprehension.

Overall, moderate consistency was found between the dependent variables studied and the outcome measure types utilised. As previously mentioned, self-report measures were the most consistent measure across all categories. For behavioural and physiological measures, there was no pattern observed, with the exception of interventions in the 'Behavioural and Attitudinal

Change' category where physiological measures were consistently absent. This finding suggests a lack of established parameters for assessing psychological outcomes, highlighting the need to understand which types of outcome measures are better equipped for measuring specific psychological domains. Establishing such parameters would enable consistent and holistic assessment of a given psychological construct or behaviour across different interventions, spanning both objective and subjective data.

Furthermore, upon a more detailed evaluation of the main categories of dependent variables, a lack of consistency in the specific assessments used to measure the same constructs was observed, especially regarding self-report measurements. This inconsistency highlights the standardisation issues discussed earlier. While having multiple tools to assess a construct can benefit measurement reliability, an excess of diverse tools may hinder comparability across studies, especially when no common assessment is widely adopted. Additionally, a moderate trend was found where researchers developed their own questionnaires. Two main types of self-designed questionnaires emerged: those created from a combination of items from multiple validates scales and those with self-created items particular to the intervention.

Although both types may raise concerns about the overall validity and reliability of the self-designed questionnaires, those with self-created items are particularly problematic when also considering generalisability. The psychology field encourages researchers to utilise validated scales because they are designed with "specific administration instructions, carefully chosen items, and a specific medium for delivery" to ensure that the results confidently assess the construct in question (Clairmont, 2022). Items in validated scales have been evaluated by experts to avoid misinterpretations, biases, and statistical issues such as ceiling and floor effects (Clairmont, 2022). Considering the importance of using validated scales to measure the extent of

a psychological construct, XR psychological interventions that rely on self-designed questionnaires and scales may compromise the validity and reliability of their findings and limit their generalisability.

Regarding behavioural measures, increased variance was found in the types of assessments used. This variability can be directly attributed to the use of XR technology, which places individuals in customised environments. Because XR environments aim to evoke specific emotions or behaviours from the established setting, the evaluation methods for these behaviours are bound to vary according to the expected responses. Due to the diversity across XR interventions, environments, and objectives, standardising behavioural measurements proves challenging given the individual differences in reactions and responses. However, efforts to standardise these outcome measures can still be made when using validated behavioural scales, such as the *Clinically Administered PTSD Scale for DSM-5 (CAPS-5)* for assessing PTSD, like the approach suggested for self-report measures to establish consistency between interventions.

Furthermore, moderate associations were observed between physiological measures and the dependent variables. Interventions studying mental health exhibited the greatest use of physiological measures, particularly centred on cardiovascular and psychophysiological measures. This finding suggests a higher applicability of these measures in assessing psychological constructs related to stress, emotional regulation, functionality, anxiety, depression, and symptom improvement in clinical populations. Consistent with existing literature, indicators such as EDA, BP, HR, HRV, and RR are known indicators of sympathetic nervous system arousal, corresponding to increases in stress, negative emotions, and anxiety (Gaffey & Wirth, 2014). Thus, the consistent use of physiological measures in these interventions targeting psychological enhancement accurately aligns with previous research and

underscores the potential use of these in providing a comprehensive understanding of mental health-related constructs, particularly within XR interventions designed to influence sympathetic nervous system activity.

Interventions focusing on social well-being ranked second in their use of physiological measures, specifically regarding eye movement data. This emphasis is supported by research highlighting the crucial role of eye gaze during natural social interactions, found to be essential for successful communication (Cañigueral & de C. Hamilton, 2019). Studies suggests that interpersonal gaze processing provides valuable insights to social behaviour, resulting in increasing attention towards the significance of eye gaze coordination between individuals for accurate signal processing and provision of meaningful feedback (Cañigueral & de C. Hamilton, 2019). Additionally, eye contact enables individuals to gather information about others' interests, emotions, and thought processes (Cañigueral & de C. Hamilton, 2019). Considering these insights into the role of eye gaze in emotion perception and connectedness, the increased use of eye gaze measurements across interventions targeting social well-being proves meaningful.

Implications of the Results

Findings and Implications for Future Research

The findings of this literature review support the continued development and refinement of XR-based psychological interventions, emphasizing the need for converging operations in outcome measures. Given the multifaceted nature of psychological constructs, obtaining a holistic assessment of mental health and cognitive improvements is crucial. This study highlights the necessity for further research to explore the integration of behavioural and physiological measures with self-report tools, particularly leveraging the advancements in XR technology. With XR technology enabling real-life simulation environments, researchers can generate real

emotions and continuously track physiological changes during interventions, offering a more comprehensive understanding of the effectiveness of these technologies for use as part of psychological treatment.

Theoretical Contributions

This literature review provides a comprehensive framework for categorising psychological interventions that utilise XR technologies across various mental health domains. The identification and holistic classification of the diverse outcome measures enables the creation of a repertoire on frequently used measures according to the dependent variable studied. This approach may aid in the standardisation of measures in future interventions and further theoretical developments. Moreover, the recommendation to synthesize outcome measures for standardisation highlights the importance of developing unified assessment tools that can better inform about the desired dependent variables.

Practical Applications

A lack of methodological rigor in the design process of XR psychological interventions, particularly with how dependent variables are measured, was identified in the review. This issue raises questions about the necessity of having numerous validated questionnaires and scales that essenstially measure the same construct. The absence of standardisation in measuring dependent variables creates ripple effects that hinder comparability in methodologies and design across studies.

To address this standardisation issue, the checklist tool was developed. The checklist aids in maintaining consistency in intervention methodologies by ensuring that researchers consider the same critical elements when designing their XR psychological intervention. This minimises

variability in study designs focusing on similar constructs and enables more reliable and generalisable findings.

Furthermore, founded in published literature, the checklist provides a wide array of dependent variables that are currently identified as important, prompting researchers to consider these insights during the design process. The use of the comprehensive repertoire of psychological constructs provided by the literature review promotes a holistic approach to evaluating specific constructs and ensures important factors are not overlooked.

Moreover, the checklist enhances the validity and reliability of their intervention designs by prompting considerations that may mitigate potential biases. By fostering critical thinking and self-reflection throughout the design process, it encourages researchers to evaluate their decisions and approaches, strengthening the justifications of their intervention design.

Ultimately, the checklist serves as a practical tool to guide researchers in the emerging field of XR psychological interventions. It facilitates intervention planning, promotes comprehensive evaluation of psychological constructs, encourages innovation in exploring new combinations of related dependent variables, and promotes adherence to the use of complementary combinations of dependent variable, as identified by previous research. Its application aims to enhance the standardisation and comparability of XR interventions across various settings and populations.

Future Research Directions

Future studies could build upon this systematic review to further explore the potential for converging operations when measuring multifaceted psychological constructs, thereby improving measurement reliability. Given the extensive number of self-report and behavioural outcome measures identified, future researchers could ensure these assessments are validated.

Additionally, synthesising and standardising these identified outcome measures is recommended for future psychological interventions by reviewing assessments measuring the same construct and removing those that lack validity or reliability. Researchers are also encouraged to continue exploring advancements in XR technologies to optimise therapeutic interventions and enhance user experience, ensuring they do not become impersonal or robotic. Given the recency in the advancements in XR technologies, it is crucial for future research to investigate the long-term effects and sustainability of XR psychological interventions on mental health outcomes.

Additionally, subsequent steps for the checklist tool developed from this systematic review involve consolidating the mind maps into a cohesive online questionnaire. A promising online platform for this would be Qualtrics XM, a recognised survey maker tool, which also allows for rapid dissemination. In principle, it would be the most beneficial for researchers to complete this online survey after identifying the fundamental elements of their XR psychological intervention, such as the main dependent variable and the targeted population, and prior to the process of intervention design.

Furthermore, ongoing plans include integrating these mind maps with those developed by the second researcher, which focus on the paterns in independent variables found across most articles considered in this review, with some exeptions given the differing inclusion and exclusion criteria employed. Additional recommendations to optimise this checklist include reviewing further literature to address the multiple-choice options that currently lack subsequent questions or insights. To enhance the comprehensiveness of the checklist, responses provided under the ‘None of the above’ multiple-choice options should be considered for future includion, after identification of supporting literature.

Limitations of Evidence Included in the Review

The limitations within the included literature have implications for this review. Due to the vast literature gatheres and time constraints, there was limited ability to comprehensively assess the quality and rigor of all interventions included and their respective study designs. This review primarly focused on the dependent variables and methods of measurement restricting detailed readings to the introduction section and outcome measures subsections of each article. While this approach may introduce biases, the main goal is not to evaluate the effectiveness of different XR treatments but rather to provide an overview of the psychological constructs studied using XR and their measurement methods.

Language barriers also posed moderate constraints on the evidence within this literature review, as non-English studies were excluded, limiting geographic diversity and the generalisability of findings beyond English-speaking popoulations. However, despite these language limitations, psychological constructs targeted by XR interventions are universally experienced, and behavioural or physiological measures not reliant on verbal communication can be employed across cultures.

While the reviewed interventions target diverse populations such as those with mental disorders, the elderly, or children, specific measures tailores to these groups may differ from those used for the general population, potentially affecting comparability. Despite limitations in generalisability, findings remain relevant within respective populations and offer frameworks for future research expansion.

Limitations of the Review Process

Despite efforts to comprehensively select, examine, and categorise the literature, several inherent limitations in the review process should be addressed. It should be noted that all literature was gathered from a single database, Scopus, which may unintentionally perpetuate publication biases as only Scopus-indexed literature was considered. Moreover, due to the broad scope of the literature review encompassing all current XR psychological interventions, the search query was developed through a deductive approach where reviewers continually added exclusion terms derived from the titles of undesired articles found in the previous search results.

Some limitations arose from working alongside but independently from another reviewer. Firstly, despite a discussion on the inclusion and exclusion criteria, their application was left to each reviewer's discretion, particularly for ambiguous studies. Secondly, a comparable limitation applied to labeling of application areas during the data extraction phase, leading to varied classification patterns between reviewers. However, recognising these inconsistencies, the application areas were considered indicative for categorisation decisions, but final categories were not based on these labels. Future reviews should consider a collaborative codebook specifying inclusion and exclusion criteria and establish standardised labels using an inductive approach to ensure consistency across reviewers and facilitating holistic view of the data.

Conclusion

In conclusion, this literature review significantly contributes to the understanding of current XR-based psychological interventions by synthesizing extensive literature, categorising studies based on their dependent variables, and exploring the diverse methods used to measure outcomes. The patterns in dependent variables utilised across the literature were used as a basis to develop seven overarching categories, namely those targeting mental health, exposure therapy, a combination of mental health and cognitive functioning, cognitive functioning alone, health, social well-being, and behavioural and attitudinal change.

The dependent variable categories identified emphasised the versatility of XR technology in assessing both subjective experiences and objective data through self-report, behavioural, and physiological measures. Self-report measures were the most prevalent across all categories, especially in interventions related to exposure therapy, mental health and cognitive functioning, and health, where every study employed this form of assessment. Behavioural assessments were most frequently used across studies assessing mental health, cognitive functioning and behavioural and attitudinal change. Physiological measures were particularly important for interventions assessing social well-being, while completely absent for those focusing on behavioural or attitudinal change.

Inconsistencies were found in the approaches of the same or similar constructs, emphasising the need for standardisation in methodologies to allow for comparability between interventions. To address this lack of methodological rigor, a checklist grounded in the literature reviewed was developed to assist researchers in developing standardised XR psychological interventions by nudging them to consider the same critical elements in their design, minimising variability in study design and measurement approaches while enhancing reliability and

generalisability. Additionally, as a secondary benefit, fostering critical thinking and self-reflection on decisions regarding intervention design. Moving forward, there is a need to refine and expand this checklist to encompass broader psychological domain, which will aid in advancing the methodological rigor of the emerging XR psychological intervention field.

References

American Psychological Association. (n.d.). *APA Dictionary of Psychology*. Dictionary.apa.org. https://dictionary.apa.org/behavioral-approach-task

American Psychology Association. (n.d.). *APA PsycNet*. Psycnet.apa.org. https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft15072-000

Antanavičius, J. (2023, March 30). *Extended reality: opportunities, success stories and challenges in health and education*. Media and Learning. https://media-and-learning.eu/type/featured-articles/extended-reality-opportunities-success-stories-and-challenges-in-health-and-education/

Banire, B., Al Thani, D., Qaraqe, M., Mansoor, B., & Makki, M. (2020). Impact of mainstream classroom setting on attention of children with autism spectrum disorder: an eye-tracking study. *Universal Access in the Information Society*, *20*(20). https://doi.org/10.1007/s10209-020-00749-0

Barillot, L., Chauvet, C., Besnier, M., Jaafari, N., Solinas, M., & Chatard, A. (2023). Effect of environmental enrichment on relapse rates in patients with severe alcohol use disorder: protocol for a randomised controlled trial. *BMJ Open*, *13*(5), e069249–e069249. https://doi.org/10.1136/bmjopen-2022-069249

Basharat, A., Mehrabi, S., Muñoz, J. E., Middleton, L. E., Cao, S., Boger, J., & Barnett-Cowan, M. (2023). Virtual reality as a tool to explore multisensory processing before and after engagement in physical activity. *Frontiers in Aging Neuroscience*, *15*, 1–22. https://doi.org/10.3389/fnagi.2023.1207651

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity.

Journal of Neuroscience Methods, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Billett, H. H. (1990). *Hemoglobin and Hematocrit*. National Library of Medicine; Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK259

Cañigueral, R., & de C. Hamilton, A. F. (2019). The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00560

Centers for Disease Control and prevention. (2022, June 3). *Body Mass Index (BMI)* . Centers for Disease Control and Prevention. https://www.cdc.gov/healthyweight/assessing/bmi/index.html

Chamberland, C., Bransi, M., Boivin, A., Jacques, S., Gagnon, J., & Tremblay, S. (2023). The effect of augmented reality on preoperative anxiety in children and adolescents: A randomized controlled trial. *Pediatric Anesthesia*, *34*(2), 153–159. https://doi.org/10.1111/pan.14793

Chen, Y., Zhou, Z., Cao, M., Liu, M., Lin, Z., Yang, W., Yang, X., Dhaidhai, D., & Xiong, P. (2022). Extended Reality (XR) and telehealth interventions for children or adolescents with autism spectrum disorder: Systematic review of qualitative and quantitative studies. *Neuroscience & Biobehavioral Reviews*, *138*, 104683. https://doi.org/10.1016/j.neubiorev.2022.104683

Clairmont, A. (2022, January 21). *Why Use Validated Measures? Part 1: What Is a Validated Measure? - EvalCorp*. Evalcorp.com. https://evalcorp.com/why-use-validated-measures-part-1/

Cleveland Clinic. (2022, January 19). *VO2 max: How To Measure and Improve It*. Cleveland Clinic. https://health.clevelandclinic.org/what-is-vo2-max-and-how-to-calculate-it

De Luca, R., Naro, A., Colucci, P. V., Pranio, F., Tardiolo, G., Billeri, L., Le Cause, M., De Domenico, C., Portaro, S., Rao, G., & Calabrò, R. S. (2021). Improvement of brain functional connectivity in autism spectrum disorder: an exploratory study on the potential use of virtual reality. *Journal of Neural Transmission, 128*. https://doi.org/10.1007/s00702-021-02321-3

De Vos, A., De Troyer, R., & Stove, C. (2019, January 1). *Chapter 57 - Biomarkers of Alcohol Misuse* (V. R. Preedy, Ed.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/abs/pii/B978012813125100057X

Diamond, L. M., & Otter-Henderson, K. D. (2009). Handbook of Research Methods in Personality Psychology. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Google Books* (pp. 370–371). Guilford Press. https://books.google.nl/books?hl=en&lr=&id=XHwS3PU6uroC&oi=fnd&pg=PA370&dq=importance+of+physiological+measures+in+psychology&ots=JEJdNYTKZE&sig=D97gAyXndTxOqj4h14cJG9e3FWM&redir_esc=y#v=onepage&q&f=false

EFP Instrumentendatabank. (n.d.). *Benton Visual Retention Test - EFP Instrumentendatabank*. Kennisdatabank.efp.nl. Retrieved May 20, 2024, from https://kennisdatabank.efp.nl/instrumenten/benton-visual-retention-test

Freeman, D., Freeman, J., Ahmed, M., Haynes, P., Beckwith, H., Rovira, A., Miguel, A. L., Ward, R., Bousfield, M., Riffoid, L., Kabir, T., Waite, F., & Rosebrock, L. (2023). Behavioural and Cognitive Psychotherapy Automated VR therapy for improving positive self-beliefs and psychological well-being in young patients with psychosis: a proof of concept evaluation of Phoenix VR self-confidence therapy. *Behavioural and Cognitive*

Psychotherapy, *52*(3), 277–287. Cambridge University Press. https://doi.org/10.1017/S1352465823000553

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *Lancet Psychiatry*, *5*(8). Lancet Psychiatry. https://doi.org/10.1016/S2215-0366(18)30226-8

Freeman, D., Rosebrock, L., Waite, F., Loe, B. S., Kabir, T., Petit, A., Dudley, R., Chapman, K., Morrison, A., O'Regan, E., Aynsworth, C., Jones, J., Murphy, E., Powling, R., Peel, H., Walker, H., Byrne, R., Freeman, J., Rovira, A., & Galal, U. (2022). Virtual reality (VR) therapy for patients with psychosis: satisfaction and side effects. *Psychological Medicine*, *53*(10), 1–12. https://doi.org/10.1017/s0033291722001167

Gaffey, A. E., & Wirth, M. M. (2014). Psychophysiological Measures. *Encyclopedia of Quality of Life and Well-Being Research*, 5181–5184. https://doi.org/10.1007/978-94-007-0753-5_2315

Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations - Why are they so highly variable? *ResearchGate*.

Gamito, P., Oliveira, J., Alves, C., Santos, N., Coelho, C., & Brito, R. (2020). Virtual Reality-Based Cognitive Stimulation to Improve Cognitive Functioning in Community Elderly: A Controlled Study. *Cyberpsychology, Behavior, and Social Networking*, *23*(3), 150–156. https://doi.org/10.1089/cyber.2019.0271

Gamito, P., Oliveira, J., Matias, M., Cunha, E., Brito, R., Lopes, P. F., & Deus, A. (2021).

Virtual Reality Cognitive Training Among Individuals With Alcohol Use Disorder Undergoing Residential Treatment: Pilot Randomized Controlled Trial. *Journal of Medical Internet Research*, *23*(1), e18482. https://doi.org/10.2196/18482

Geraets, C. N. W., van der Stouwe, E. C. D., Pot-Kolder, R., & Veling, W. (2021). Advances in immersive virtual reality interventions for mental disorders: A new reality? *Current Opinion in Psychology*, *41*(1), 40–45. https://doi.org/10.1016/j.copsyc.2021.02.004

Gualtieri, C. T., & Johnson, L. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*(7), 623–643. https://doi.org/10.1016/j.acn.2006.05.007

Ingram, K. M., Espelage, D. L., Merrin, G. J., Valido, A., Heinhorst, J., & Joyce, M. (2019). Evaluation of a virtual reality enhanced bullying prevention curriculum pilot trial. *Journal of Adolescence*, *71*, 72–83. https://doi.org/10.1016/j.adolescence.2018.12.006

Lahiri, U., Warren, Z., & Sarkar, N. (2011). Design of a Gaze-Sensitive Virtual Social Interactive System for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(4), 443–452. https://doi.org/10.1109/TNSRE.2011.2153874

Lee, H. (2020). A Conceptual Model of Immersive Experience in Extended Reality . *PsyArXiv Preprints*, 1–23. https://doi.org/10.31234/osf.io/sefkh

Levy, A. N., Nittas, V., & Wray, T. B. (2023). Patient Perceptions of In Vivo Versus Virtual Reality Exposures for the Treatment of Anxiety Disorders: Cross-Sectional Survey Study. *JMIR Formative Research*, *7*(1), e47443. https://doi.org/10.2196/47443

Lin, Y., Zhang, W. J., & Watson, L. G. (2001). On Proposal of Function-Behavior-State

Framework as Refinement of EID Framework of Human-Machine Interface Design. *Human Friendly Mechatronics*, 61–66. ScienceDirect. https://doi.org/0.1016/B978-044450649-8/50012-7

Mahdieh Sasaninezhad, Moradi, A., Sharareh Farahimanesh, Mohammad Hasan Choobin, & Mostafa Almasi-Dooghaee. (2024). Enhancing cognitive flexibility and working memory in individuals with mild cognitive impairment: Exploring the impact of virtual reality on daily life activities. *Geriatric Nursing*, *56*, 32–39. https://doi.org/10.1016/j.gerinurse.2023.12.008

Mayo Clinic. (2022, March 19). *Electrocardiogram (ECG or EKG) - Mayo Clinic*. Mayoclinic.org; Mayo Clinic. https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983

MedlinePlus. (2022, July 7). *Cortisol test: MedlinePlus lab test information*. Medlineplus.gov. https://medlineplus.gov/lab-tests/cortisol-test/

Meins, I. A., Muijsson-Bouwman, D. C., Nijman, S. A., Greaves-Lord, K., Veling, W., Pijnenborg, G. H. M., & van der Stouwe, E. C. D. (2023). VR-SOAP, a modular virtual reality treatment for improving social activities and participation of young people with psychosis: a study protocol for a single-blind multi-centre randomized controlled trial. *Trials*, *24*(1). https://doi.org/10.1186/s13063-023-07241-z

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Nahum, M., Lee, H., Fisher, M., Green, M. F., Hooker, C. I., Ventura, J., Jordan, J. T., Rose, A.,

Kim, S.-J., Haut, K. M., Merzenich, M. M., & Vinogradov, S. (2020). Online Social Cognition Training in Schizophrenia: A Double-Blind, Randomized, Controlled Multi-Site Clinical Trial. *Schizophrenia Bulletin*, *47*(1), 108–117. https://doi.org/10.1093/schbul/sbaa085

Nater, U. M., & Rohleder, N. (2009). Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: Current state of research. *Psychoneuroendocrinology*, *34*(4), 486–496. https://doi.org/10.1016/j.psyneuen.2009.01.014

National Collaborating Centre for Mental Health (UK). (2011). PSYCHOLOGICAL AND PSYCHOSOCIAL INTERVENTIONS. In *www.ncbi.nlm.nih.gov*. British Psychological Society (UK). https://www.ncbi.nlm.nih.gov/books/NBK65484/

Ortega, J., Plaska, C. R., Gomes, B. A., & Ellmore, T. M. (2022). Spontaneous Eye Blink Rate During the Working Memory Delay Period Predicts Task Accuracy. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.788231

Pennington, D. L., Reavis, J. V., Cano, M. T., Walker, E., & Batki, S. L. (2022). The Impact of Exercise and Virtual Reality Executive Function Training on Cognition Among Heavy Drinking Veterans With Traumatic Brain Injury: A Pilot Feasibility Study. *Frontiers in Behavioral Neuroscience*, *16*. https://doi.org/10.3389/fnbeh.2022.802711

Pratama, N. R., Wafa, I. A., Budi, D. S., Putra, M., Wardhana, M. P., & Wungu, C. D. K. (2022). PRISMA 2020 checklist. *Plos.figshare.com*. https://doi.org/10.1371/journal.pone.0261350.s001

Price, P. C., Jhangiani, R., & Cuttler, I-Chant. A. (n.d.). 4.1 Understanding Psychological Measurement. In *Research Methods in Psychology*. PressBooks. https://opentext.wsu.edu/carriecuttler/chapter/understanding-psychological-measurement/

Rayyan Help Center. (2024, February 9). *Understanding Rayyan: A Comprehensive Overview*. Rayyan: Intelligent Systematic Review. https://help.rayyan.ai/hc/en-us/articles/22697630697617-Understanding-Rayyan-A-Comprehensive-Overview

Ritsert, F., Elgendi, M., Galli, V., & Menon, C. (2022). Heart and Breathing Rate Variations as Biomarkers for Anxiety Detection. *Bioengineering (Basel, Switzerland)*, *9*(11), 711. https://doi.org/10.3390/bioengineering9110711

Ryser, V.-A. (2023). Measuring Psychological constructs. FORS Guide No. 22, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. doi:10.24449/FG-2023- 00022

Salazar, L. F., Parrott, D. J., DiLillo, D., Gervais, S., Schipani-McLaughlin, A. M., Leone, R., Swartout, K., Simpson, L., Moore, R., Wilson, T., Flowers, N., Church, H., & Baildon, A. E. (2019). Study protocol for a randomized controlled trial of RealConsent2.0: a web-based intervention to promote prosocial alcohol-involved bystander behavior in young men. *Trials*, *24*(1), 1–16. ResearchGate. https://doi.org/10.1186/s13063-023-07797-w

Shen, Z.-H., Liu, M.-H., Wu, Y., Lin, Q.-Q., & Wang, Y.-G. (2022). Virtual-reality-based social cognition and interaction training for patients with schizophrenia: A preliminary efficacy study. *Frontiers in Psychiatry*, *13*. https://doi.org/10.3389/fpsyt.2022.1022278

Smith, P. G., Morrow, R. H., & Ross, D. A. (2015, June). *Types of intervention and their development*. Nih.gov; OUP Oxford. https://www.ncbi.nlm.nih.gov/books/NBK305514/

Stone, V., Baron-Cohen, S., O'riordan, M., Jones, R., & Plaisted, K. (1999). *Faux Pas Recognition Test Faux Pas Recognition Test (Child Version)*. https://docs.autismresearchcentre.com/tests/FauxPas_Child.pdf

Sun, Y., Zhou, H., Liu, C., Wang, A., Yue, C., & Zhang, M. (2022). Sound-induced flash illusion

is modulated by the depth of auditory stimuli: Evidence from younger and older adults. *Attention, Perception & Psychophysics*, *84*(6), 2040–2050. https://doi.org/10.3758/s13414-022-02537-9

Tam, J. W., & Schmitter-Edgecombe, M. (2013). The Role of Processing Speed in the Brief Visuospatial Memory Test – Revised. *The Clinical Neuropsychologist*, *27*(6), 962–972. https://doi.org/10.1080/13854046.2013.797500

Tsang, M. M. Y., & Man, D. W. K. (2013). A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. *Schizophrenia Research*, *144*(1-3), 51–62. https://doi.org/10.1016/j.schres.2012.12.024

Uman, L. S. (2011). Systematic Reviews and Meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie Canadienne de Psychiatrie de l'Enfant et de L'adolescent*, *20*(1), 57–59. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024725/

Uomori, K., & Yolanda, M. (1994). Binocular Eye Movements While Showing a Moving Stimulus to One Eye Only. *Studies in Visual Information Processing*, *5*, 169–183. ScienceDirect. https://doi.org/10.1016/B978-0-444-81808-9.50021-4

Uttl, B., & Graf, P. (1997). Color-Word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology*, *19*(3), 405–420. https://doi.org/10.1080/01688639708403869

Venkatasubramanian, G., & Danivas, V. (2013). Current perspectives on chlorpromazine equivalents: Comparing apples and oranges! *Indian Journal of Psychiatry*, *55*(2), 207. https://doi.org/10.4103/0019-5545.111475

Vickers, J. N. (1995, January 1). *Gaze Control in Basketball Foul Shooting* (J. M. Findlay, R.

Walker, & R. W. Kentridge, Eds.). ScienceDirect; North-Holland. https://www.sciencedirect.com/science/article/pii/S0926907X05800443

Whitfield, J. B. (2001). Gamma Glutamyl Transferase. *Critical Reviews in Clinical Laboratory Sciences*, *38*(4), 263–355. https://doi.org/10.1080/20014091084227

World Health Organization. (2022). *Mental disorders*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/mental-disorders

World Health Organization (WHO). (2024). Psychological interventions implementation manual: integrating evidence-based psychological interventions into existing services. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.

XRHealth. (n.d.). *▷ Virtual Reality Exposure Therapy | XRHealth*. XR Health. Retrieved June 25, 2024, from https://www.xr.health/virtual-reality-exposure-therapy/

Yu, K., Wang, L., Lv, S., Ye, X., Liu, L., Zheng, X., Jin, R., Zhou, D., Zhang, Y., Min, G., & Wu, S. (2023). Using functional near-infrared spectroscopy to study effects of virtual reality intervention for adolescents with depression in a clinical setting in China: study protocol for a prospective, randomised, controlled trial. *BMJ Open*, *13*(12), e074129. https://doi.org/10.1136/bmjopen-2023-074129

Appendix A

Article Distribution According to their Year of Publication

Year of Publication	Final Search Query on Scopus (N= 612)	Included studies (N= 167)
2024	10	5
2023	98	27
2022	86	27
2021	62	24
2020	60	15
2019	53	14
2018	40	9
2017	36	12
2016	25	5
2015	28	6
2014	24	7
2013	25	9
2012	14	1
2011	13	9
2010	6	1
2009	5	1
2008	3	0
2007	4	0
2006	4	1
2005	1	0
2004	5	0
2003	1	0
2002	2	1
2001	3	0
2000	2	2
1999	1	0
1998	1	0

Appendix B

Inclusion and Exclusion Keywords Specified in Rayyan

Inclusion

Virtual reality (n= 450); VR (n= 271); randomised trial (n= 130); controlled trial (n= 87); randomized controlled trial (n= 71); Augmented reality (n= 21); RCT (n= 16); randomised controlled trial (n= 11); AR (n= 10); ER (n= 5); Extended reality (n= 1);

Exclusion

Animal (n= 4); case report (n= 3); rodents (n= 2); animals (n= 1); rodent (n= 1); mice (n= 1); rats (n= 1); rat (n= 1).

Appendix C

Reasons for Article Exclusion in Rayyan

1. No use of Extended Reality (XR) in study (n= 65)
2. Use of XR for medical purposes, not psychological (n= 39)
3. Not a psychological intervention/ No psychological construct is studied (n=36)
4. Research using XR, not an intervention (n=22)
5. Use of XR but not as part of the intervention (n=14)
6. Evaluation of current state of XR use in psychological research (n=11)
7. Assessment of an intervention, not an intervention (n=10)
8. VR used only in combination with another treatment/intervention type (n=8)
9. Proposal for future XR research (n=6)
10. No abstract (n=5)
11. Full paper not publicly available online (n=5)
12. Summary of recent psychological research with no XR use (n=4)
13. Paper only available in a foreign language (n= 4)
14. Case study (n=4)
15. Use of XR for cognitive assessment and/or diagnosis (n=4)
16. Conference proceedings (n=3)
17. Research for future XR psychological intervention (n=3)
18. Use of AI chatbot, not classified as XR (n=2)
19. Position paper (n=1)
20. Editorial (n=1)
21. Duplicated article (n=1)

Appendix D

Reasons for Article Exclusion in the Data Extraction Table

1. Full article could not retrieved (n =36)
2. For introduction use only (n =24)
3. Medical intervention/ all participants have a medical condition (n =22)
4. Non-immersive use of VR (n =8)
5. Study protocol (n =8)
6. Not a psychological intervention (n =6)
7. Research using VR/ VR not used as part of the intervention/ No relevant use of VR (n =6)
8. Case study (n =3)
9. Not available in English (n =3)
10. Explorative of VR experiences (n =2)
11. Control group (no VR) is not comparable to Experimental group (VR) (n =1)
12. Evaluation of an intervention (n =1)
13. Using VR only in combination with other treatment/intervention type (n =1)

Appendix E

Application Areas: Descriptions and Distribution Across Studies

E1. Descriptions

No.	Application Area	Description
1	Addiction treatment	Encompassed interventions targeting any addictive behaviour or addictive substance (e.g., alcohol, gambling, drugs).
2	Behaviour change/ Attitude change	Included interventions focused on increasing prosocial behaviour, decreasing antisocial behaviour, and altering attitudes.
3	Cognitive intervention	Included interventions targeting executive control functions such as attention, memory, problem-solving, and decision-making skills.
4	Development of framework	Comprised studies aimed at guiding the structuring of future interventions. Studies in this category were not considered for inclusion.
5	Eating behaviour	Encompassed interventions addressing relationships with food, body image attitudes related to weight and size, as well as eating disorders and food perception.
6	Education purposes	Included interventions aiming to teach specific skills or enhance knowledge about a particular subject among participants.
7	Elderly care	Encompassed studies aimed at enhancing psychological or cognitive functioning specifically in older adults.
8	Enhancing psychological well-being	Comprised interventions targeting various aspects of well-being and quality of life, including mood regulation, restorativeness, self-confidence, and stress management.
9	Enhancing social functioning	Encompassed interventions aimed at enhancing social and emotional intelligence, which includes developing skills in recognizing emotions and reducing social anxiety.

No.	Application Area	Description
10	Environment research	Consists of interventions that examine the impact of physical environments on cognitive functioning or psychological well-being. However, as the focus of this literature review does not include changes in stimuli, this category, developed by the second reviewer, is not applicable.
11	Exposure therapy	Encompassed studies utilising virtual reality (VR) exposure therapy as the intervention method.
12	Generalisation of VR interventions	Included studies that applied previously established interventions in novel environments or with different populations.
13	Health	Encompassed interventions focused on improving physical well-being, including promoting exercise adherence and improving sleep quality.
14	Medical	Comprised interventions associated with medical contexts, including procedures (e.g., childbirth), tasks (e.g., intravenous insertion success) or conditions (e.g., tinnitus).
15	Mental disorder treatment	Included interventions aimed at alleviating symptoms of a specific mental disorder or enhancing skills within those disorders.
16	Mental health	Comprised interventions specifically addressing aspects of mental health such as anxiety and depression. This category contrasts with the "Enhancing psychological well-being" category, which focuses on broader areas of psychological wellness.
17	Palliative care	Included in this category are interventions tailored for patients undergoing palliative care.
18	PTSD	Encompassed interventions aimed at reducing symptoms associated with PTSD.
19	Public speaking	Included interventions focused on improving public speaking skills.

No.	Application Area	Description
20	Research for future VR interventions	Comprised of studies presenting compiled research to inform the development of future VR interventions. Studies in this category were not considered for inclusion.
21	Safety	Included interventions aiming to enhance awareness and safety across various domains, including occupational safety and operating room (OR) safety.
22	Not Applicable	N/A

E2. Frequencies of Studies by Application Area for "All Studies" and "Included Studies"

No.	Application Area	All studies (N= 289)	Included studies (N= 167)
1	Addiction treatment	6	7
2	Behaviour change/ Attitude change	21	4
3	Cognitive intervention	12	6
4	Development of framework	1	N/A
5	Eating behaviour	10	9
6	Education purposes	6	5
7	Elderly care	8	9
8	Enhancing psychological well-being	40	25
9	Enhancing social functioning	15	25
10	Environment research	7	N/A
11	Exposure therapy	26	14
12	Generalisation of VR interventions	3	N/A
13	Health	11	7
14	Medical	20	4
15	Mental disorder treatment	46	20
16	Mental health	23	19
17	Palliative care	2	1
18	PTSD	7	9
19	Public speaking	1	2
20	Research for future VR interventions	16	N/A
21	Safety	4	1
22	Not Applicable	5	N/A

Note. As indicated in the table above, some application areas are more frequent under the "Included Studies" column. This is due to a revision of the application areas once the studies were classified as "included" and examined in more depth.

Appendix F

Mental Health: Dependent Variables & Outcome Measures

F1. Enhancing Psychological Well-being: Stress and Coping

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Brief Virtual Reality Mindfulness is More Effective than Audio Mindfulness and Colouring in Reducing Stress in University Students	Stress	One item on prior experience with mindfulness; World Health Organisation Well-being Index (WHO-5); Perceived Stress Scale (PSS-10); One item on perceived engagement with the procedure.	N/A	Heart rate
2	Flourishing-Life-Of-Wish Virtual Reality Relaxation Therapy (FLOW-VRT-Relaxation) Outperforms Traditional Relaxation Therapy in Palliative Care: Results from a Randomized Controlled Trial	End-of-life care distress	The Chinese version of the Edmonton Symptom Assessment System (CESAS).	N/A	N/A
3	Effects of Virtual Reality Natural Experiences on Factory Workers' Psychological and Physiological Stress	Stress	The Four-Dimensional Symptom Questionnaire (4DSQ); The Positive and Negative Affect Scales (PANAS); The Perceived Stress Scale (PSS).	N/A	Blood pressure (BP)–systolic BP (SBP) and diastolic BP (DBP), heart rate variability (HRV), electrocardiogram (ECG).
4	"Room to Reflect": A Pilot Workplace Resiliency Intervention for Nurses	Resilience	Connor-Davidson Resilience Scale (CD-RISC 10); Self-designed questionnaire assessing ease of use, accessibility, and effectiveness*.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Efficacy of a Mindfulness-based Programme with and Without Virtual Reality Support to Reduce Stress in University Students: A Randomized Controlled Trial	Stress	Perceived Stress Scale (PSS); The State-Trait Anxiety Inventory (STAI); Positive and Negative Affect Schedule (PANAS); Utrecht Work Engagement Survey Scale-Students (UWES-S); Maslach Burnout Inventory-Student Survey (MBI-SS); Emotional Regulation Questionnaire (ERQ); Mechanistic measures used include: Five Facet Mindfulness Questionnaire (FFMQ); The Self-Compassion Scale (SCS).	N/A	N/A

F2. Enhancing Psychological Well-being: Mood States and Emotional Regulation

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Decreased Emotional Dysregulation Following Multi-Modal Motion-Assisted Memory Desensitization and Re-consolidation Therapy (3MDR): Identifying Possible Driving Factors in Remediation of Treatment-Resistant PTSD	Emotional regulation and emotional dysregulation in PTSD	Difficulties in Emotion Regulation Scale-18 (DERS-18)	N/A	N/A
2	Virtual Reality and Music's Impact on Psychological Well-being	Regulation of mood states and perceived presence	Abbreviated Profile of Mood States Questionnaire (POMS); Spatial Presence Experience Scale (SPEC)*.	N/A	N/A
3	EARTH of Well-Being System: A Pilot Study of an Information and Communication Technology-based Positive Psychology Intervention	Mood regulation and emotional well-being	State and Trait Anxiety Inventory (STAI); Beck Depression Inventory II (BDI-II); Self-designed questionnaire assessing acceptance and perceived usefulness; Self-designed Mood Scale assessing subjective change in mood after intervention.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Moving Beyond Role-Play: Evaluating the Use of Virtual Reality to Teach Emotion Regulation for the Prevention of Adolescent Risk Behavior Within a Randomized Pilot Trial	Emotion regulation and risk behaviour	Affect Dysregulation Scale; Difficulties in Emotion Regulation Scale (DERS-18), specifically the Lack of Emotional Awareness and Limited Access to Emotion Regulation Strategies sub-scales; Emotional Self-Efficacy Scale; 10-item Sexual Risk Knowledge quiz; Efficacy for Sexual Risk Prevention scale; Self-Efficacy for Condom Use scale	N/A	N/A
5	EEG-Based VR Scene Adaptive Generation System for Regulating Emotion	Emotion regulation	N/A	N/A	Self-assessment Model (SAM), EEG, Heart Rate (HR), and Galvanic Skin Reaction (GSR).
6	Borderline Personality Traits and Emotion Regulation Strategies in Adolescents: The Role of Implicit Theories	Emotion dysregulation	Borderline Personality Features Scale for Children (BPFS-C); Implicit Theories of Emotion Scale; Cognitive Emotion Regulation Questionnaire (CERQ); One item concerning a hypothetical treatment choice between (1) a form of psychological intervention, (2) medication, (3) a combination of medication and psychological intervention or (4) no treatment other than standard monitoring to help cope with emotional difficulties.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
7	Virtual Reality to Evaluate the Impact of Colorful Interventions and Nature Elements on Spontaneous Walking Gaze, and Emotion	Affective states	Feeling and Affect Table to Score Arousal and Appraisal (FAtAAL); Self-declared affective experiences were collected post-trials.	Spontaneous walking speed	Photoplethysmo-graphic (PPG) data, gaze behaviour, heart rate (HR), temperature, electrodermal activity (EDA).
8	Augmented Experiences: Investigating the Feasibility of Virtual Reality as part of a Workplace Well-Being Intervention	Affect	Positive and Negative Affect Schedule (PANAS).	Stroop test	Heart rate
9	Improving the Emotional Health of Juveniles with VR Game	Emotional well-being	Self-designed questionnaire assessing emotional state, engagement level, and adherence to the game. Self-designed scales assessing self-perceived leadership skills, delegation skills, decision prioritisation, and stress management skills.	Player interactions with non-player characters (NPCs) within the game were observed.	N/A

F3. Enhancing Psychological Well-being: Quality of Life and Functionality

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effects of Virtual Reality Sessions on the Quality of Life, Happiness, and Functional Fitness among the Older People: A Randomized Controlled Trial from Taiwan	Quality of Life (QoL), happiness, and functional fitness	EuroQol 5 dimensions VAS (EQ-5D-3L) Questionnaire; Mini version of the Chinese Happiness Inventory (mini-CHI).	Single-leg standing test to assess static balance; Back scratch test to evaluate shoulder joint flexibility.	Weight, body-mass index (BMI), waist circumference (WC), waist-to-hip ratio, diastolic blood pressure (BP; mmHg) and systolic BP (mmHg).
2	Different Types of Virtual Natural Environments Enhance Subjective Vitality Through Restorativeness	Restorativeness and subjective vitality	Seven items from the Perceived Restorativeness Scale (PRS-11); Subjective Vitality Scale.	N/A	N/A

F4. Enhancing Psychological Well-being: Self-Perception and Agency

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Effect of Sense of Agency on Self-Efficacy Beliefs: A Virtual Reality Paradigm	Sense of agency and self-efficacy	One item assessing sense of agency, "To which extent did you feel you had control over the 3D hand throughout the previous block of trials?"; One item assessing self-efficacy, "To which extent are you certain that you can conduct this task successfully?"; Levenson LoC questionnaire; Body ownership (BOw) questionnaire	N/A	N/A
2	Brief Report: Improving Employment Interview Self-efficacy Among Adults with Autism and Other Developmental Disabilities Using Virtual Interactive Training Agents (ViTA)	Self-efficacy and confidence	VITA-DMF Self-Efficacy Scale.	Marino Interview Assessment Scale (MIAS)	N/A
3	A Single-Session VR Intervention Addressing Self-compassion and Self-criticism With and Without Perspective Change: Results of a Randomized Controlled Experiment	Self-criticism and self-compassion	Screening measures included: Forms of Self-Criticizing/Attacking and Self-Reassuring Scale (FSCRS). Outcome measures included: Self-Compassion and Self-Criticism Scale (SCCS); Visual Analogue Scale (VAS) to assess momentary positive affect, negative affect, self-compassion and self-esteem; Igroup Presence Questionnaire (IPQ)*	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Physical Versus Virtual Reality–Based Calm Rooms for Psychiatric Inpatients: Quasi-Randomized Trial	Self-reported well-being and physiological arousal	Visual Analogue Scale (VAS) to assess self-perceived well-being.	N/A	Blood pressure (BP) and heart rate (HR).
5	Development and Validation of a Virtual Reality-based Training Program for Promoting Subjective Well-being	Subjective well-being	Mental Health Continuum-Short Form (MHC-SF); Rosenberg's Self-Esteem Scale (RSES); Dispositional Hope Scale (DHS); Life Orientation Test Revised–Korean version (LOT-R).	N/A	N/A
6	Automated VR Therapy for Improving Positive Self-beliefs and Psychological Well-being in Young Patients with Psychosis: A Proof of Concept Evaluation of Phoenix VR Self-Confidence Therapy	Positive self-beliefs and psychological well-being	Positive Self Scale (OxPos), Brief Core Schema Scale (BCSS); Warwick-Edinburgh Well-being Scale (WEMWBS); Client Satisfaction Questionnaire (CSQ); Oxford-VR Side Effects Checklist (O-VRSE)*; Adverse events were recorded, view Appendix XXX.	N/A	N/A
7	Self-Confidence and Paranoia: An Experimental Study Using an Immersive Virtual Reality Social Situation	Paranoia severity	Paranoid Thoughts Scale Part B (GPTS-B); Visual analogue scale (VAS) assessing confidence; Social Comparison Scale (SCS); State Social Paranoia Scale (SSPS).	N/A	N/A

F5. Enhancing Psychological Well-being: Psychological Well-being in Older Adults

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Effects of Therapeutic Virtual Reality Experience to Promote Mental Well-being in Older People Living with Physical Disabilities in Long-term Care Facilities	Mental well-being, depressive symptoms, loneliness, health-related quality of life, and perceived social support	World Health Organization Five Well-being Index (WHO-5); Patient Health Questionnaire (PHQ-9); The Chinese version of the 6-item De Jong Gierveld Loneliness Scale (DJGLS); The Hong Kong version of the EuroQol 5-dimensions instrument with a five-level scale (EQ-5D-5L); Multiple Scale of Perceived Social Support (MSPSS).	N/A	N/A
2	Affective Out-World Experience via Virtual Reality for Older Adults Living with Mild Cognitive Impairments or Mild Dementia	Emotion regulation and mental health	Visual Analogue Scale (VAS) assessing emotional state; Slater-Usoh-Steed Questionnaire (SUS)*; System Usability Scale (SUS)*.	Observed Emotion Rating Scale (OERS); Semi-structured interviews to evaluate technology acceptance.	Eye-tracking data
3	Combination of 3-Dimensional Virtual Reality and Hands-on Aromatherapy in Improving Institutionalized Older Adults' Psychological Health: Quasi-experimental Study	Perceived stress, happiness, sleep quality, meditation experience, and life satisfaction	Oxford Happiness Inventory; Perceived Stress Scale (PSS); Pittsburgh Sleep Quality Index (PSQI); Shortened version of the 10-item Experiences During Meditation (EOM-DM) scale	N/A	N/A
4	Effect of Sensorimotor Rehabilitation Based on an Immersive Virtual Reality Model on Mental Health	Global mental health: depression, anxiety	RYFF Well-being Scales; YESAVAGE scale also known as Geriatric Depression Scale (GDS); Goldberg General health questionnaire (GHQ-12), specifically the anxiety dimension.	N/A	N/A

F6. Enhancing Psychological Well-being: Mental Health in Extreme Environments

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	A 7-Day Space Habitat Simulated Task: Using a Projection-Based Natural Environment to Improve Psychological Health in Short-Term Isolation Confinement	Negative mental states	The State-Trait Anxiety Inventory (STAI-Y1 and STAI-Y2) Scales; The Positive and Negative Affect Schedule (PANAS).	N/A	N/A
2	Well-being in Isolation: Exploring Artistic Immersive Virtual Environments in a Simulated Lunar Habitat to Alleviate Asthenia Symptoms	Stress	Shortened version of the Singapore Mental Wellbeing Scale (SMWEB-S); Pictoral Self-Assessment Mannequin (SAM) questionnaire; Self-designed multi-dimensional feedback tool to evaluate participants' responses to presented virtual experiences; Self-designed Short Presence Scale (SPS); The Self-Assessment Manikin (SAM); Self-designed Short well-being scale; Self-designed report assessing symptoms of asthenia.	N/A	Heart rate (HR) and electrodermal activity (EDA).

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Mobile Health–Supported Virtual Reality and Group Problem Management Plus: Protocol for a Cluster Randomized Trial Among Urban Refugee and Displaced Youth in Kampala, Uganda	Mental health outcomes	Patient Health Questionnaire-9 (PHQ-9) for depression; Modified depression literacy scale validated in LMICs; Inventory of Attitudes Towards Seeking Mental Health Services; Kidcope and Self-Compassion Scale for Youth; Brief Version of the Internalized Stigma of Mental Illness scale; WHO-Five Well-being Scale (WHO-5); WHO Disability Assessment Schedule.	N/A	N/A
4	Natural Scene Virtual Reality as a Behavioral Health Countermeasure in Isolated, Confined, and Extreme Environments: Three Isolated, Confined, Extreme Analog Case Studies	Psychological well-being	Initial two case studies measures included: Positive Affect, Negative Affect Schedule (PANAS) assessing mood pre and post VR; Modified Reality Judgement and Presence Questionnaire (MRJPQ)*; Value of VR questionnaire*. Final study measures included: MRJPQ*; Perceived Restorativeness Scale (PRS-11); Profile of Mood States (POMS) to assess more subtle mood changes than the PANAS.	Semi-structured exit interviews assessed utility of the VR programmes, possible improvements, and perceived value and acceptability of natural scene VR.	N/A

F7. Anxiety and Depression: Anxiety

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	A Virtual Psychiatric Ward for Orientating Patients Admitted for the First Time	Anxiety and understanding of psychiatric wards	Six-item Chinese version of the State–Trait Anxiety Inventory (STAI) questionnaire; IBM Computer System Usability Questionnaire (IBM SCUQ)*.	Level of understanding test (LUT)	Heart rate variability (HRV)
2	Exposure to Virtual Social Interactions in the Treatment of Social Anxiety Disorder: A Randomized Controlled Trial	Social fears and social anxiety	Screening measures included: Social Interaction Anxiety Scale (SIAS); Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I). Outcome measures included: Liebowitz Social Anxiety Scale-Self Report (LSAS-SR); Fear of Negative Evaluation Scale-Brief Form (FNE-B); Depression Anxiety Stress Scale (DASS-21); Personality Disorder Belief Questionnaire (PDBQ), specifically assessing Avoidant personality disorder; Eurohis Quality of Life Scale (EUROHIS-QOL 8-item index).	5 min impromptu speech to evaluate levels of behavioural avoidance, measuring speech duration and speech performance.	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Treating Childhood Social Anxiety Disorder With Virtual Environments and Serious Games: A Randomized Trial	Social anxiety	Social Phobia and Anxiety Inventory for Children (SPAI-C); Parental reports obtained included: Child Behavior Checklist (CBCL). Clinician ratings included: Children's Global Assessment Scale (C-GAS); Clinical Global Impressions Scale (CGI).	Parental reports obtained included: Social Phobia and Anxiety Inventory for Children-Parent Version (SPAIC-PV); Clinician-Administered PTSD Scale (CAPS); Structured Clinical Interview for DSM-IV (SCID I and II); Self-monitoring and reporting of PTSD symptoms (daily) including nightmares, flashbacks, total hours of sleep, anger, physical rage episodes, verbal rage episodes, and anxiety.	N/A
4	Understanding the Psycho-Physiological Implications of Interaction With a Virtual Reality-Based System in Adolescents With Autism: A Feasibility Study	Social anxiety levels	Screening measures included: Childhood Autism Rating Scale 2nd Edition—High Functioning (CARS 2-HF).	N/A	Physiological signals, e.g., pulse plethysmogram (PPG), skin temperature (SKT), and electrodermal activity (EDA)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Design of a Physiology-Sensitive VR-Based Social Communication Platform for Children With Autism	Social anxiety	Screening measures included: Social Responsiveness Scale (SRS). Outcome measures included: Social Communication Questionnaire (SCQ); Spence Children's Anxiety Scale (SCAS) completed by their caregiver.	N/A	Pulse plethysmogram (PPG), skin temperature (SKT), and electrodermal activity (EDA). Physiological indexes were fed to fuzzy-logic based Intelligent Anxiety Predictor (IAP) units for predicting one's anxiety level.
6	Virtual Reality Acceptance and Commitment Therapy Intervention for Social and Public Speaking Anxiety: A Randomized Controlled Trial	Social and public speaking anxiety	Social Interaction Anxiety Scale (SIAS); Personal Report of Communication Apprehension (PRCA-24); Mental Health Continuum (MHC-SF); Perceived Stress Scale (PSS); Visual Analog Scales (VAS) assessing comfortability, stress, nervousness, and willingness to give a speech; The Comprehensive Assessment of ACT Processes (CompACT); Self Compassion Scale – Short Form (SCS-SF); Brief Fear of Negative Evaluation Revised Scale (BFNE).	N/A	N/A
7	An Experimental Study on Fear of Public Speaking using a Virtual Environment	Public speaking anxiety	Personal Report of Confidence as a Public Speaker (PRCS); Self-designed questionnaire assessing subjective experience to the virtual environment (VE)*.	Researcher observed and listened to the presentation from a remote location.	Heart rate

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
8	Feasibility of Autism-focused Public Speech Training Using a Simple Virtual Audience for Autism Spectrum Disorder	Public speech stress and self-confidence	Autism Spectrum Quotient – Japanese Version (AQ-J); Attention Deficit Hyperactivity Disorder Rating Scale (ADHD-RS).	Liebowitz Social Anxiety Scale (LSAS)	Salivary cortisol levels
9	The Effect of Virtual Reality Spiritual Emotional Freedom Technique *VR-SEFT) Therapy on Anxiety and Cortisol in Drug Patients in Makassar Class I State Detention Center	Anxiety	Hamilton Rating Scale for Anxiety (HRS-A)	N/A	Cortisol levels using ELISA method
10	Art Making and Virtual Reality: A Comparison Study of Physiological and Psychological Outcomes	Anxiety, stress, and mood	Positive and negative affect (PANAS); State-Trait Anxiety Inventory (STAI); Perceived Stress Scale (PSS).	N/A	Heart rate, electrodermal activity, saliva collection to measure Salivary Alpha Amylase (sAA) enzyme quantity.
11	Virtual Tool for the Development of Strategies to Reduce the Stress in Students	Stress and anxiety	Cohen's Perceived Stress Scale (PSS) adapted and validated in Spanish.	N/A	N/A
12	Effects of Restorative Environment and Presence on Anxiety and Depression Based on Interactive Virtual Reality Scenarios	Anxiety and depression states	Restoration Environment Scale (RES); Presence Questionnaire (PQ)*; State-Trait Anxiety Inventory (STAI); Self-rating Depression Scale (SDS).	N/A	Electroencephalogram (EEG); Electromyography (EMG).

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
13	Daily Exposure to Virtual Nature Reduces Symptoms of Anxiety in College Students	Anxiety, depression and rumination	Penn State Worry Questionnaire (PSWQ); Mood and Anxiety Symptom Questionnaire (MASQ); Rumination Reflection Questionnaire (RRQ); One item assessing sleep quantity: "how much sleep did you get on average during the past week?"; Godin-Shephard Leisure-Time Physical Activity Questionnaire. Possible moderator measured: Engagement with Beauty Scale.	N/A	N/A
14	The Effect of Virtual Reality and Buzzy on First Insertion Success, Procedure-Related Fear, Anxiety, and Pain in Children during Intravenous Insertion in the Pediatric Emergency Unit: A Randomized Controlled Trial	Anxiety and fear	Wong-Baker Faces Pain Rating Scale (WBFPRS); Color Analog Scale (CAS); Children's Anxiety Meter-State (CAMS), and Child Fear Scale (CFS).	Emotional Appearance Scale for Children (EASC); Difficult Intravenous Access (DIVA) score to evaluate first-attempt intravenous insertion success.	Vital signs (Body temperature, Heart rate, Respiratory rate, Blood pressure)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
15	Comparison of the Effectiveness of Virtual Reality and Chewing Mint Gum on Labor Pain and Anxiety: A Randomized Controlled Trial	Childbirth anxiety and pain	Fertility questionnaire; Pain visual analogue; Spielberger Anxiety Questionnaire.	N/A	Routine delivery unit care measures were recorded: "analgesia, oxytocin levels, misoprostol, pethidine, hyoscine, promethazine and atropine (except for the use of Entonox and spinal and epidural anesthesia, control fetal heart rate, vaginal exams, vital signs recording, etc)."
16	Virtual Reality on Mobile Phones to Reduce Anxiety in Outpatient Surgery	Anxiety	Visual Analogue Scale assessing anxiety (VAS-A).	N/A	N/A
17	The Effect of Augmented Reality on Preoperative Anxiety in Children and Adolescents: A Randomized Controlled Trial	Pre-operative anxiety	Yale Preoperative Anxiety Scale (mYPAS-SF)	N/A	N/A
18	The Use of Pre-Operative Virtual Reality to Reduce Anxiety in Women Undergoing Gynecological Surgeries: A Prospective Cohort Study	Pre-operative anxiety	State-Trait Anxiety Inventory (STAI); Hospital Anxiety and Depression Scale (HADS); EuroQol 5-dimension 3-level (EQ-5D-3L); Numerical Rating Scale (NRS) for pain.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
19	Effect of Immersive Virtual Reality Education before Chest Radiography on Anxiety and Distress among Pediatric Patients: A Randomized Clinical Trial	Anxiety and distress	N/A	Observational Scale of Behavioral Distress (OSBD) scale; The time taken for each child's radiographic procedure (from entering the radiography room to image production) was recorded. Followingly, the radiology technologist rated the difficulty of the procedure for each child using a Numerical Rating Scale.	N/A
20	An Innovative Virtual Reality Experience in the PICU: A Pilot Study	Perceived calming effect, enjoyment and satisfaction with VR	Self-designed questionnaire (completed by parents) assessing (1) whether the child enjoyed using VR, (2) whether they enjoyed watching their child use VR, (3) whether the child wanted to extend the VR time, (4) whether the VR calmed their child, (5) whether VR was confusing, difficult, or uncomfortable. Self-designed questionnaire (for parents) on prior VR use, adverse effects, and possible future use.	Research investigator monitored participant anxiety or distress through verbal cues or body language.	N/A

F8. Anxiety and Depression: Depression

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Using Functional Near-Infrared Spectroscopy to Study Effects of Virtual Reality Intervention for Adolescents with Depression in a Clinical Setting in China: Study Protocol for a Prospective, Randomised, Controlled Trial	Depression, anxiety, and executive function	Hamilton Anxiety Scale (HAMA); Beck Depression Inventory-II (BDI-II); Hamilton Depression Scale-17 (HAMD-17); Treatment Emergent Symptom Scale (TESS).	Verbal Fluency Task (VFT)	Haemoglobin changes measured by f(NIRS)
2	Does Virtual Reality Increase the Efficacy of Psychotherapy for Young Adults with Mild-to-Moderate Depression? A Study Protocol for a Multi-center Randomized Clinical Trial	Severity of depression and distress	Beck Depression Inventory II (BDI-II); Clinical Outcomes in Routing Evaluation-Outcome Measure (CORE-OM); CORE-SFB; Depression and Anxiety, and Stress Scales (DASS-21); Session Rating Scale 3.0 (SRS 3.0); The Mini International Neuropsychiatric Intervire (MINI). Consumer Reports Effectivenes Scale (CRES-4); Life Satisfaction Scale.	Client change interview	N/A
3	Psychedelic Replications in Virtual Reality and Their Potential as a Therapeutic Instrument: An Open-Label Feasibility Study	Depressive symptoms	Emotional State Questionnaire (EST-Q2); The MontgomeryÅsberg Depression Scale (MADRS); Psychological Insight Questionnaire (PIQ); Ego-Dissolution Inventory (EDI); The Revised Mystical Experience Questionnaire (MEQ30).	N/A	N/A

F9. Symptom Management in Mental Disorders: Psychotic Disorder Symptoms

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The CHALLENGE Trial: The Effects of a Virtual Reality-Assisted Exposure Therapy for Persistent Auditory Hallucinations Versus Supportive Counselling in People with Psychosis: Study Protocol for a Randomised Clinical Trial	Distress of auditory hallucinations, perceived voice power, patient acceptance of voices, patient's ability to respond to voices in an assertive way, social and daily functioning.	Psychotic Symptoms Rating Scales (PSYRATS-AH); PSYRATS-AH-Frequency; PSYRATS-AH-Distress; Voice Power Differential Scale; Voices Acceptance and Action Scale, specifically the VAAS-Action subscale; The Personal and Social Performance Scale (PSP).	Assertive Responding to Voices	N/A
2	Virtual Reality Therapy for Refractory Auditory Verbal Hallucinations in Schizophrenia: A Pilot Clinical Trial	Auditory verbal hallucinations, depression, and QoL	Psychotic Symptoms Rating Scale (PSYRATS); Beliefs About Voices Questionnaire-Revised (BAVQ-R); Positive And Negative Syndrome Scale (PANSS); Beck Depression Inventory-II (BDI-II); Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form.	N/A	N/A
3	Can Virtual Reality Reduce Reality Distortion? Impact of Performance Feedback on Symptom Change in Schizophrenia Patients	Delusion severity	Paranoia, Obsessive-compulsive and Depression questionnaire (POD).	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Virtual Reality (VR) Therapy for Patients with Psychosis: Satisfaction and side effects	Psychosis symptoms	Four items from the Client Satisfaction Questionnaire (CSQ)*; Two items on "number of sessions and experience of the staff member supporting VR delivery"; Oxford – VR Side Effects Checklist (O-VRSE); Oxford Agoraphobic Avoidance Scale (O-AS); Revised Green et al. Paranoid Thoughts Scale (R-GPTS); Mobility Inventory for Agoraphobia – Alone subscale (MIA); Paranoia Worries Questionnaire; Patient Health Questionnaire (PHQ-9); Questionnaire about the Process of Recovery; Beck Hopelessness Scale (BHS); Oxford Cognitions and Defences Questionnaire (O-CDQ); Negative Voices When Outside.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Integrating a Virtual Reality Relaxation Clinic Within Acute Psychiatric Services: A Pilot Study	Psychological well-being	Six Visual Analogues Scales (VAS) assessing current self-reported levels of subjective (1) stress, (2) relaxation, (3) anxiety, (4) sadness, (5) happiness, (6) connectedness to nature. Two VAS assessing VR helpfulness and immersiveness.	Thematic analyses were conducted using process notes from two independent researchers and therapy session transcripts to understand therapists' experiences, challenges, and suggested improvements in delivering VR relaxation. Data on violent incidents and restrictive practices on the wards were extracted from databases.	N/A

F9. Symptom Management in Mental Disorders: Post-Traumatic Stress Disorder Symptoms

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Use of Virtual Reality as a Tool to Overcome the Post-Traumatic Stress Disorder of Pensioners	PTSD symptoms	Brief Fear of Negative Evaluation Scale (BFNE); Beck Anxiety Inventory (BAI); PTSD Checklist for the DSM-5 (PCL-5) .	Trail Making Test (TMT)	N/A
2	The Efficacy of Trauma Management Therapy: A Controlled Pilot Investigation of a Three-Week Intensive Outpatient Program for Combat-Related PTSD	PTSD Symptoms, social isolation, anger and depression	Clinician—Administered PTSD Scale (CAPS); PTSD Checklist Military Version (PCL-M); Structured Clinical Interview for DSM-IV (SCID I and II); Quality of Life Inventory (QoLI); Medical Outcome Study Short Form-36 Health Survey (SF-36); The Behaviour Rating Inventory of Executive Function– Adult Version (BRIEF-A); Trauma-Related Guilt Inventory (TRGI); Clinical Global Impressions Scale (CGI); Hamilton Rating Scale for Depression (HAMD); Clinician Administered Rating Form of Functional Indicators.	Self-monitoring of PTSD symptoms (daily) including nightmares, flashbacks, total hours of sleep, anger, physical rage episodes, verbal rage episodes, and anxiety.	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Moving Forward from Moral Injury: A Mixed Methods Study Investigating the Use of 3MDR for Treatment Resistant PTSD.	Presence of moral injury	The Military Injury Symptom Scale–Military Short Form (MISS-M-SF); PTSD Checklist for DSM-5 (PLC-5); Patient Health Questionnaire (PHQ-9); Generalised Anxiety Disorder Scale (GAD-7), Outcome Questionnaire 45 (OQ-45); the Dissociative Experiences Questionnaire (PDE-Q), the Connor-Davidson Resilience Scale (CD-RISC-25); the Alcohol Use Disorder Identification Test (AUDIT); the Difficulties in Emotion Regulation Scale (DERS-18).	Clinically Administered PTSD Scale for DMS-5 (CAPS-5)	N/A
4	Interactive Motion-Assisted Exposure Therapy for Veterans with Treatment-Resistant Post-traumatic Stress Disorder: A Randomized Controlled Trial	PTSD symptom severity	Life Events Checklist for DSM-5; Mini-International Neuropsychiatric Interview-Plus for DSM-IV; PTSD Checklist for the DSM-5 (PCL-5); Hospital Anxiety and Depression Scale (HADS); Post-traumatic Avoidance Behaviour Questionnaire (PABQ); Interpersonal Support Evaluation List; The validated Cantril's Ladder of Life.	Clinician-Administered PTSD Scale for DSM-5 (CAPS-5); Emory Treatment Resistance Interview for PTSD	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Quantitative Changes in Mental Health Measures with 3MDR Treatment for Canadian Military Members and Veterans	Treatment-resistant PTSD symptoms	PTSD Checklist for DSM-5 (PCL-5); The Military Injury Symptom Scale–Military Short Form (MISS-M-SF); The Patient Health Questionnaire-9 (PHQ-9); Generalized Anxiety Disorder Scale-7 (GAD-7); Peritraumatic Dissociation Event Questionnaires (PDEQ); Alcohol Use Disorder Identification Test (AUDIT); The Difficulties in Emotion Regulation (DERS-18) Scale; Outcomes Questionnaire-45 (OQ-45); Connor Davidson Resilience Scale (CD-RISC).	Clinician-Administered PTSD Scale, Fifth Edition (CAPS-5) for DSM-5	N/A
6	Veteran Perceptions of Virtual Reality to Assess and Treat Post-traumatic Stress Disorder	PTSD symptoms	Subjective Units of Distress (SUDS) on a scale of 0-100; 7-item Presence Questionnaire (PQ)*; Self-designed questionnaire assessing level of absorption experienced during VR (e.g., "How much did the visual aspects of the environment engage you?")*.	Clinician-Administered PTSD Scale (CAPS)	N/A
7	A Randomized, Head-to-Head Study of Virtual Reality Exposure Therapy for Post-traumatic Stress Disorder	PTSD symptoms	N/A	Clinician-Administered PTSD Scale (CAPS)	N/A
8	Investigating Relationships Between PTSD Symptom Clusters Within Virtual Reality Exposure Therapy for OEF/OIF Veterans	PTSD symptoms	PTSD Symptom Scale (PSS).	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
9	A Virtual Reality Paradigm as an Analogue to Real-life Trauma: Its Effectiveness Compared with the Trauma Film Paradigm	Intrusion frequency and self-rated vividness and emotionality of recalled memories.	Eysenck Personality Questionnaire (EPQ), specifically the Neuroticism trait; State-Trait Anxiety Inventory (STAI); Visual Analogue Scales (VAS) assessing mood; Self-designed questionnaire evaluating 4 statements about the film from 0-100 (e.g. "I felt personally involved" and "The events were unpredictable") from 0-100. Intrusions were recorded using a paper-and-pencil intrusion diary for 7 days after watching the film. Diary compliance was measured by rating the statement "I was often unable (or often forgot) to report my intrusions in the diary" on a 10-point scale.	Memory vividness and emotionality was assessed by asking participants to recall and visualise the moment from the film or VR scene that they considered most unpleasant for 10 seconds and "then rate its vividness and emotionality on two VAS that ranged from 0 (not vivid/unpleasant) at all to 100 (extremely vivid/unpleasant)".	N/A
10	Psycho-physiologic Identification of Sub-threshold PTSD in Combat Veterans	Fear in PTSD	Checklist-Military version (PLC-M), Short Form Survey (SF-35); Patient Health Questionnaire (PLQ); Generalised Anxiety Disorder (GAD-7).	Clinician-Administered PTSD Scale (CAPS); Fear potentiated state paradigm validated at Emory University; Physical examination by a medical professional with a detailed assessment of balance and smell.	Electromyography (EMG), galvanic skin response (GSR), electrocardiogram (EGC), respiratory rate (RR).

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
11	A Controlled Trial for PTSD in Mexican Victims of Criminal Violence	PTSD symptoms	Posttraumatic Stress Symptom Scale, Self-Report (PSS-SR); State-Trait Anxiety Inventory (STAI); Beck Depression Inventory (BDI); Treatment satisfaction questionnaire [adapted]*.	Clinician Administered PTSD Scale (CAPS-1)	N/A

F10. Symptom Management in Mental Disorders: Autism Spectrum Disorder (ASD) Symptoms

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effect of Cognitive Training Based on Virtual Reality on the Children with Autism Spectrum Disorder	Severity of Autism Spectrum Disorder (ASD) symptoms	Autism Behavior Checklist (ABC), Childhood Autism Rating Scale (CARS).	Clancy Autism Behavior Scale (CABS)	N/A
2	Multi-modal Job Interview Simulator for Training of Autistic Individuals	Job interview outcomes and perceived stress levels	N/A	Utterances spoken during the interviews were labeled and classified using a slightly modified dialogue act annotation scheme described in (Chakravarty et al., 2019; Farzana et al., 2020). Six self-designed labels were also included and four were adapted from (Jurafsky and Shriberg, 1997; Stolcke et al., 2000); Behaviour analyst trained in rating autistic behaviour gave ground truth stress ratings from the interview recordings; Facial expressions were assessed using a webcam and the FaceAPI software which captured and processed facial images every 5.5 seconds, returning a confidence score for each of the eight universally recognized emotions.	Photoplethysmo-graphic (PPG) data, electrodermal activity (EDA).

F11. Symptom Management in Mental Disorders: Substance Use Disorder (SUD) Symptoms

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effect of Environmental Enrichment on Relapse Rates in Patients with Severe Alcohol Use Disorder: Protocol for a Randomised Controlled Trial	Rate of relapse to alcohol use	Craving measures included the Obsessive Compulsive Drinking Scale Questionnaire; The Craving Induction Protocol of Fox et al. using a Visual Analogue Scale.	Standard Implicit Association Test; An adapted version of Moeller et al.'s probabilistic image choice task to assess the seeking for alcohol-related stimuli; Soleymani et al.'s Visual Research Test for alcohol-related stimuli utilising eye tracking; Proportion of individuals who relapsed 2 weeks after treatment– relapse defined as drinking more than 5 drinks per occasion or more than 5 times a week. Mid-term relapse assessed at 1 month and 3 months.	Breath via breathalyser and a blood test for the measurements of carbohydrate deficient tranferrin (CDT) and gamma-glutamyl transpeptidase (GGT).
2	Virtual Reality Cognitive Training Among Individuals with Alcohol Use Disorder Undergoing Residential Treatment: Pilot Randomized Controlled Trial	Cognitive functioning	Montreal Cognitive Assessment (MoCA); Frontal Assessment Battery (FAB).	Rey Complex Figure (RCF); Toulouse Pierón test (TP); Wisconsin Card Sorting Test (WCST).	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Gambling Exposure in Virtual Reality and Modification of Urge to Gamble	Urge to gamble	Canadian Problem Gambling Index (CPGI) including additional items regarding the most recent video lottery terminal (VLT) gambling experience and alcohol and tobacco habits while gambling; Gambling inventory including the aforementioned VLT items and novel items evaluating the realism of the virtual gambling environment, whether the gambling environment is similar to the environment in which they most often gamble, and whether they experienced any distress with regard to their urge to gamble after exposure; Self-designed questionnaire assessing urge to gamble on VLTs and perceived self-efficacy.	N/A	N/A
4	Effects of a Virtual Reality-Based Motivational Reinforcement + desensitization intervention program on psychological craving and addiction memory in female MA-dependent young adults	Craving and addiction memory	VAS scale assessing subjective psychological craving for drugs; Addiction Memory Intensity Scale (AMIS).	N/A	Heart rate (HR) and blood pressure (BP).

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Individual Variables Related to Craving Reduction in Cue Exposure Treatment	Cigarette craving	VAS scale assessing; Fagerstrom Test for Nicotine Dependence (FTND); Nicotine Dependence Syndrome Scale (NDSS); Minnesota Nicotine Withdrawal Scale (MNWS); State-Trait Anxiety Inventory (STAI); Beck Depression Inventory-Second Edition (BDI-II).	Delay Discounting (DD) task	N/A
6	Assessing the Attentional Bias of Smokers in a Virtual Reality Anti-Saccade Task Using Eye Tracking	Craving, attentional bias (AB) and inhibitory control	Craving measures included: verbal rating from 0 to 100; Visual Analog Scale (VAS)). Fagerström Test for Nicotine Dependence (FTND); Brief Questionnaire of Smoking Urges (QSU). One item assessing presence ("How strongly, from 0-100, do you agree with the following statement: In the computer-generated world, I had the impression of being there.")*; Igroup Presence Questionnaire (IPQ)*; Virtual Reality Sickness Questionnaire (VRSQ)*.	N/A	Binocular eye movement and respiration.

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
7	Avatar-Assisted Therapy: A Proof-of-Concept Pilot Study of a Novel Technology-based Intervention to Treat Substance Use Disorders	Treatment success	N/A	Screening measures included a standardised treatment intake interview. Outcome measures included: Treatment completion (attendance to 16 sessions); Urine drug screen-positive samples measured at baseline and randomly during treatment duration; Rearrest information measured weekly during treatment duration.	N/A

Appendix G

Exposure Therapy: Dependent Variables & Outcome Measures

G1. Fear of Specific Phobias

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effectiveness of Self-Guided App-Based Virtual Reality Cognitive Behavior Therapy for Acrophobia: A Randomized Clinical Trial	Acrophobia	Acrophobia Questionnaire (AQ); AQ-Anxiety; Attitudes Toward Heights Inventory (ATHI); Beck Anxiety Inventory (BAI); Patient Health Questionnaire–9 item (PHQ-9); System Usability Scale (SUS)*; Igroup Presence Questionnaire (IPQ)*.	N/A	N/A
2	Automated Psychological Therapy Using Immersive Virtual Reality for Treatment of Fear of Heights: A Single-blind, Parallel-group, Randomised Controlled Trial	Acrophobia	Heights Interpretation Questionnaire (HIQ); Acrophobia Questionnaire (AQ); Improving Access to Psychological Therapies (IAPT) phobia scale; Adverse events were recorded, view Appendix O.	N/A	N/A
3	Effectiveness of a Smartphone-based, Augmented Reality Exposure App to Reduce Fear of Spiders in Real-Life: A Randomized Controlled Trial	Arachnophobia	Subjective Units of Distress Scale (SUDS) in a Behavioural Approach Test (BAT); Beck Depression Inventory II (BDI-II); Fear of Spiders Questionnaire (FSQ); Augmented Reality Immersion (ARI) questionnaire*.	Approach behaviour towards spiders with a BAT test	N/A
4	Virtual Reality Treatment of Flying Phobia	Aviophobia	Anxiety Disorders Interview Schedule (ADIS-IV); Subjective Units of Discomfort Scale (SUDS); Fear Record (FR); Maladjustment Scales (MS); Danger Expectations and Flying Anxiety Scales (DEFAS); Fear of Flying Questionnaire (FFQ).	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	A Cognitive and Virtual Reality Treatment Program for the Fear of Flying	Aviophobia	Questionnaire d'Evaluation de la Peur de l'Avion (QEPA) "composed with the Flight Anxiety Situations (FAS) and the Flight Anxiety Modality (FAM) in their French version".	N/A	N/A
6	Virtual Reality Exposure Therapy for Fear of Driving: Analysis of Clinical Characteristics, Physiological Response and Sense of Presence	Vehophobia	Mini International Neuropsychiatric Interview (MINI); Beck Depression Inventory (BDI); Hamilton Anxiety Scale (HAMA); State-Trait Anxiety Inventory (STAI); Driving Cognitions Questionnaire (DCQ); The Medical Outcomes Survey SF-36 (MOS SF-36); Subjective Units of Distress Scale (SUDS); Igroup Presence Questionnaire (IPQ)*.	Structured Clinical Interview for DSM-IV Axis II (SCID-II)	Heart rate
7	The Use of Mobile-Assisted Virtual Reality in Fear of Darkness Therapy	Nyctophobia	Fear Thermometer questionnaire.	N/A	N/A
8	Virtual Reality in the Treatment of Claustrophobic fear: A Controlled, Multiple-baseline Design	Claustrophobic	Subjective Units of Discomfort Scale (SUDS); Fear Record (FR); Problem-Related Impairment Questionnaire; Anxiety Sensitivity Index (ASI).	Admission interview using screening questions for each of the DSM-IV diagnostic criteria for anxiety disorders; Behavioural Avoidance Test (BAT);	N/A

G2. Fear Medical Procedure

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Virtual Reality Exposure Therapy for the Treatment of Dental Phobia: A Controlled Feasibility Study	Dentophobia	Modified Dental Anxiety Scale (MDAS); Phobia Checklist; Visual Analogue Scale-Anxiety (VAS-A); Dental Fear Scale (DFS); Subjective Unit of Distress Scale (SUDS).	Behavioural Avoidance Test (BAT)	Heart rate
2	Effects of Virtual Reality Exposure Therapy on Dentophobia in Clients of Dental Offices in Isfahan, Tehran, and Shahrekord (Iran)	Dentophobia	Modified Dental Anxiety Scale (MDAS)	N/A	N/A
3	Development of a Virtual Reality Exposure Tool as Psychological Preparation for Elective Pediatric Day Care Surgery: Methodological Approach for a Randomized Controlled Trial	Pre-operative anxiety	Yale Preoperative Anxiety Scale (mYPAS); Visual Analogue Scale (VAS) assessing anxiety levels; State-Trait Anxiety Inventory (STAI); Pain measures included: Revised Faces Pain Scale (FPS-r); Face, Legs, Activity, Cry, and Consolability (FLACC) scale; The Parents' Postoperative Pain Measure (PPPM); Pediatric Anesthesia Emergency Delirium (PAED); Child trait anxiety was assessed by parents with the Child Behavior Checklist (CBCL); Parental trait anxiety was self-reported using the trait anxiety form (20 items) of the STAI.	N/A	N/A

G3. Other Phobias

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Effectiveness of Self-Guided Virtual-Reality Exposure Therapy for Public-Speaking Anxiety	Public speaking anxiety	Public Speaking Anxiety scale (PSAS); Personal Report of Confidence as a Speaker (PRCS-SF); Brief Fear of Negative Evaluation revised scale (FNE-B); Subjective Units of Distress Scale (SUDS).	Liebowits Social Anxiety Scale (LSAS)	Heart rate
2	Exposure and Response Prevention in Virtual Reality for Patients with Contamination-Related Obsessive–Compulsive Disorder: A Case Series	Distress in contamination-related obsessive compulsive disorder (C-OCD) who primarly experience feelings of disgust	Yale-Brown Obsessive Compulsive Scale (Y-BOCS); Distress and arousal assessed using SUDS; Sense of presence*; Simulator Sickness Questionnaire*.	N/A	Heart rate, electrodermal activity
3	Trauma Management Therapy with Virtual-Reality Augmented Exposure Therapy for Combat-Related PTSD: A Randomized Controlled Trial	Combat-related PTSD: depression, anger, and social isolation	PTSD Checklist Military Version (PCL-M); Miller-Forensic Assessment of Symptoms Test (M-FAST); Clinical Global Impressions Scale (CGI); Hamilton Rating Scale for Depression (HAMD); Hamilton Rating Scale for Anxiety (HAMA); Three self-designed "treatment credibility scales were used to assess for potential differences in outcome expectancy based on group assignment."	Clini-cian—Administered PTSD Scale (CAPS); Structured Clinical Interview for DSM-IV (SCID I and II); Behavioural ratings were assessed via a daily log reporting on (a) nightmares– total hours of sleep, and severity of anger, (b) socialisation– amount of time spent in social activities.	N/A

Appendix H

Mental Health and Cognitive Functioning (MH & CF): Dependent Variables & Outcome Measures

H1. General Population

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Enhancing cognitive flexibility and working memory in individuals with mild cognitive impairment: Exploring the impact of virtual reality on daily life activities	Cognitive flexibility, working memory, daily living performance, anxiety and depression symptoms	The Lawton-Brody Instrumental Activities of Daily Living (IADL) Scale; Geriatric Depression Scale (GDS); Geriatric Anxiety Inventory (GAI).	Wisconsin Card Sorting Test (WCST-64); The Symbol Span Subtest of the Wechsler Memory Scale 4th Edition (WMS-IV); Digit Span Subtest of the Wechsler Adults Intelligence Scale 4th Edition (WAIS-IV);	N/A
2	A Short Bout of Exercise With and Without an Immersive Virtual Reality Game Can Reduce Stress and Anxiety in Adolescents: A Pilot Randomized Controlled Trial	Self-reported stress, anxiety and cognitive performance in adolescents during stressful exam weeks	Trail Making Test (TMT) Parts A and B; The Pediatric Anxiety Short Form 8a (PASF); The Psychological Stress Experiences-Short Form 8a (PSES).	N/A	Heart rate (HR)
3	The Effects of Acute Virtual Reality Exergaming on Mood and Executive Function: Exploratory Crossover Trial	Mood and executive function	Self-rating of Perceived Exertion (RPE); Japanese versions of the Profile of Mood States second edition (POMS2); Two-Dimensional Mood Scale (TDMS).	Color-Word Stroop Test (CWST)	Heart rate, spontaneous eye blink rate (sEBR) measured as a noninvasive brain dopaminergic system indicator.

H2. Elderly Population

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effect of Immersive Virtual Reality Reminiscence versus Traditional Reminiscence Therapy on Cognitive Function and Psychological Well-being among Older Adults in Assisted Living Facilities: A randomized controlled trial	Cognitive functioning and psychological well-being	Mini-Mental State Examination (MMSE) translated into the Arabic language; Revised Riff's Psychological Well-being Scale (RPWBS); CyberSickness in Virtual Reality Questionnaire (CSQ-VR)*; Experience in Immersive Virtual Environment Questionnaire*.	N/A	N/A
2	Virtual Reality Intervention to Improve Apathy in Residential Aged Care: Protocol for a multi-site non-randomised controlled trial	Apathy, cognition and depression	Apathy Evaluation Scale (AES) clinician version; Addenbrooke's Cognitive Examination III (ACE-III); Geriatric Depression Scale (GDS) Short-Form. Exploratory outcomes possibly relevant include: The Quality of Life in Alzehimer's Disease (QOL-AD) and The Revised UCLA Three-Item Loneliness Scale.	Activity levels (Actigraphy) measured using a GENEActiv wristband (placed for 48 hours at baseline and is repeated for 48 hours at the end of intervention after follow-up).	Galvanic skin response, heart rate variability
3	Using Virtual Reality to Improve Apathy in Residential Aged Care: Mixed methods study	Verbal fluency and apathy	Apathy Evaluation Scale (AES) Self-Rated Version; Slater-Usoh-Steed Presence Questionnaire (SUS)*.	Phonemic task for verbal fluency, "Name as many words as possible starting with the letter F or P". Semantic task for verbal fluency, "List as many words possible in the category of either 'animals' or 'fruits/vegetables'".	N/A

Appendix I

Cognitive Functioning: Dependent Variables & Outcome Measures

I1. Enhancing Cognitive Performance and Attention

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Restorative Effects of Short-Term Exposure to Nature in Immersive Virtual Environments (IVEs) as Evidenced by Participants' Brain Activities	Cognitive performance and self-perceived stress	Perceived Stress Scale-14 (PS-14).	Stroop task	EEG
2	Effects of Biophillic Interventions in Office on Stress Reaction and Cognitive Function: A Randomised Crossover Study in Virtual Reality	Selective attention, creativity, reaction time	N/A	Stroop color-word test for selective attention; Guilfor's Alternative Uses Test (AU test)	Blood pressure (BP), heart rate variability (HRV), electrodermal activity, eye tracking (saccades, fixations, and velocity of eye movement)
3	Virtual reality promotes greater improvements than video-stimulation screen on perceptual-cognitive skills in young soccer athletes	Decision-making, visual search behaviour, and inhibitory control performance	N/A	Passing [of the ball] decision making-performance was assessed by two experienced soccer coaches blinded to the experiment and trained to use the instrument for coding, each decision was coded as either 1 for appropriate or 0 for inappropriate. Inhibitory control and selective attention measures included: Stroop task	Eye tracking (number and duration of the fixations were obtained to examine visual search behavior during SSGs)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	A Novel Neurofeedback Attentional Enhancement Approach Based on Virtual Reality	Attention enhancement	N/A	N/A	EEG

I2. Cognitive Health in Aging

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Effectiveness of a Virtual Reality-Based Intervention on Cognitive Functions in Older Adults with Mild Cognitive Impairment: A Single-Blind, Randomized Controlled Trial	Executive functioning	Loewenstein Occupational Therapy Cognitive Assessment-Geriatric (LOTCA-G).	N/A	N/A
2	Virtual Reality-Based Cognitive Stimulation to Improve Cognitive Functioning in Community Elderly: A Controlled Study	Cognitive functioning	Montreal Cognitive Assessment (MoCA); Frontal Assessment Battery (FAB); Geriatric Depression Scale-15 (GDS-15); Satisfaction with Life Scale (SWLS), translated and adapted to Portuguese.	Rey Complex Figure (RCF)	N/A

I3. Cognitive Performance in Clinical Populations: Attention-Deficit Hyperactivity Disorder (ADHD)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Training inhibitory control in adolescents with elevated attention deficit hyperactivity disorder traits: a randomised controlled trial of the Alfi Virtual Reality programme	Response inhibition	Screening measures included: Conners-3 Parent Rating Scale; Kaufman Brief Intelligence Test–Second Edition. Outcome measures included: Attention network task; Strengths and Weaknesses of ADHD and Normal Behaviour (SWAN); Digit span (DST); Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency (UPPS-P); Social Skills Questionnaires; Assessment of Quality of Life; Social Responsiveness Scale - Second Edition; Family Assessment Device; Spence Children's Anxiety Scale (SCAS); Child Behavior Checklist (CBCL).	Balloon Analogue Risk Task -Youth; Digit span (DST); Behavior Rating Inventory Executive Function, 2nd edition (BRIEF-2); Perth Emotion Regulation Competency Inventory; NEPSY 2nd Edition (NEPSY-II); Response inhibition assessed using an anticipated response stop-signal task. Reactive response inhibition assessed by "determining the stop-signal response time (SSRT) using the integration method". Proactive response inhibition "calculated by comparing the Go reaction time between the block containing only Go trials and the blocks comprising both Stop and Go trials."	N/A
2	The Influence of State and Trait Anxiety on the Achievement of a Virtual Reality Continuous Performance Test in Children and Adolescents with ADHD Symptoms	Sustained attention, vigilance, and response inhibition.	Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV); The State-Trait Anxiety Inventory for Children (STAI-C)	Aula Nespora VR-based continuous performance test: "Omissions, comissions, response time, and motor activity (motion sensors recording all motor activity of the participant) are measured".	N/A

I4. Cognitive Performance in Clinical Populations: Autism Spectrum Disorder (ASD)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Improvement of Brain Functional Connectivity in Autism Spectrum Disorder: An Exploratory Study on the Potential Use of Virtual Reality	Cognitive functions	Self-designed psychometric battery including: Raven's Colored Progressive Matrices (RCPM); Raven's Standard Progressive Matrices (RSPM); Beck Anxiety Inventory for Youth (BAI-Y).	Modified little bell test (MTCM); Developmental Test of Visual-Motor Integration (VMI)	EEG recording
4	Mechanism of Action for Obtaining Job Offers with Virtual Reality Job Interview Training	Obtaining a job offer	Screening measures to confirm mood disorder diagnosis, PTSD diagnosis, schizophrenia diagnosis, or ASD diagnosis included: Mini-International Neuropsychiatric Interview (MINI); Social Responsivity Scale (SRS); Outcome measures included: the Repeatable Battery for the Assessment of Neuropsychological Status total score; Self-reported number of months since prior employment; Six month follow-up on whether they had received a job offer since completing the study.	Screening measures to confirm mood disorder diagnosis, PTSD diagnosis, schizophrenia diagnosis, or ASD diagnosis included: Structured Clinical Interview for DSM-IV. Outcome measures included: The Virtual Reality Job Interview Training (VR-JIT) scores, ranging from 0 to 100, assessing trainee performance on eight learning goals; Number of completed virtual interviews as an indicator of training exposure. Pretest and posttest mock job interviews with professional actors scored across nine domains, with the scores averaged to provide a single performance score for each time point.	N/A

I5. Cognitive Performance in Clinical Populations: Schizophrenia

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Virtual reality-based theory of mind intervention in schizophrenia: Preliminary efficacy results	Theory of Mind (ToM)	Positive and Negative Syndrome Scale (PANSS); Wechsler Adult Intelligence Scale (WAIS); Repeatable Battery for the Assessment of Neuropsychological Status (RBANS); Lancashire Quality of Life Profile (LQoLP).	Wisconsin Card Sorting Test (WCST-64). ToM measures included: Baron-Cohen Minds in the Eyes Test (RMET); Faux pas test; Cartoon stories task. Communicational pragmatic skills measures included: The Hungarian Metaphor and Irony test.	EEG recording
6	A Virtual Reality-based Vocational Training System (VRVTS) for People With Schizophrenia in Vocational Rehabilitation	Executive functioning	Self-designed questionnaire assessing knowledge and skills in performing sales-related activities and in self-efficacy in performing sales-related activities.	Brief Neuropsychological Cognitive Examination (BNCE); Digit Vigilance Test (DVT); Rivermead Behavioural Memory Test (RNMT); Wisconsin Card Sorting Test (WCST-CV4); Vocational Cognitive Rating Scale (VCRS).	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
7	Virtual Reality Job interview Training and 6-month Employment Outcomes for Individuals with Schizophrenia Seeking Employment	Interviewing self-confidence	Self-designed report on vocational history; Scale for the Assessment of Positive Symptoms; Scale for the Assessment of Negative Symptoms; Ratings and notes from the Mini-International Neuropsychiatric Interview if data was already available from participants who completed prior studies; Repeatable Battery for the Assessment of Neuropsychological Status; Bell-Lysaker Emotion Recognition Task; TEQ assessing VR-JIT usability, enjoyability, helpfulness, and success in instilling confidence; Self-designed questionnaire assessing self-confidence using a 7-point Likert scale.	Emotional perspective-taking task not described; Job interview role-plays scored across "nine domains: (1) conveying oneself as a hard worker, (2) sounding easy to work with, (3) conveying that one behaves professionally, (4) negotiation skills, (5) sharing things in a positive way, (6) sounding honest, (7) sounding interested in the position, (8) comfort level, (9) establishing rapport"; Participant VR-JIT performance scores were recorded (0-100), number of completed truals, and time spent performing virtual interviews.	N/A

I6. Cognitive Performance in Clinical Populations: Post-Traumatic Stress Disorder (PTSD)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
8	Using Multi-sensory Virtual Reality Nature Immersion as a Therapeutic Modality for Improving HRV and Cognitive Functions in Post-Traumatic Stress Disorder: A Pilot-Study	Cognitive functions, psychological symptoms of PTSD and subjective distress	Peritraumatic Dissociative Experiences Questionnaire (PDEQ); PTSD Checklist-5 (PCL-5); Dissociative Experiences Scale (DES-II); Patient Health Questionnaire (PHQ-9); CNS Vital Signs Cognitive Test.	Clinician-Administered PTSD (CAPS-5) scale for DSM5	Heart rate variability (HRV)
9	The Impact of Exercise and Virtual Reality Executive Function Training on Cognition Among Heavy Drinking Veterans With Traumatic Brain Injury: A Pilot Feasibility Study	Cognitive function, alcohol use, alcohol craving	Obsessive Compulsive Drinking Scale (OCDS); The Alcohol Use Disorders Identification Test (AUDIT); Beck Depression Inventory (BDI-II); Beck Anxiety Inventory (BAI); Neurobehavioural Symptom Inventory (NSI); Physical Activity Readiness Questionnaire for Everyone (PAR-Q+); Physical Activity Readiness Medical Examination (PARmed-X); Goden Leisure-time Exercise Questionnaire (GLTEQ).	Time Line Follow Back (TLFB) interview; Substance Use Disorders sections of the Structured Clinical Interview for DSM-5; Hopkins Verbal Learning Test– Revised (HVLT-R); Brief Visuospatial Memory Test– revised (BVMT-R); Wechsler Test of Adult Reading; WAIS-IV Arithmetic and Digit Span; D-KEFS Trail Making Test Condition 1; D-KEFS Trail Making Condition 4 (Letter-Number Sequencing); Design Condition 3 (Switching Dots); D-KEFS Color-Word Interference Test Condition 3; D-KEFS Color-Word Interference Test Condition 4; D-KEFS Trail Making Test, Condition 2 (Number Sequencing); D-KEFS Trail Making Test, Condition 3 (Letter Sequencing); D-KEFS Color-Word Interference Test Condition 1 (Color Naming); D-KEFS Color-Word Interference Test Condition 2 (Word Naming); DKEFS Design Fluency Condition 1 (Filled Dots); DKEFS Design Fluency Condition 2 (Empty Dots); The Reliable Digit Span (RDS).	N/A

Appendix J

Health: Dependent Variables & Outcome Measures

J1. Encouraging Physical Activity: Physical Activity and Exercise Adherence

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Virtual Reality as a Tool to Explore Multi-sensory processing before and after engagement in physical activity	Multi-sensory processing among older adults	Audiovisual response time, Sound-Induced Flash illusion (SIFI)	Temporal Order Judgements (TOJ)	N/A
2	Using Virtual Agents and Activity Monitors to Autonomously Track and Assess Self-Determined Physical Activity among Young Children: A 6-Week Feasibility Field Study	Self-determined physical activity	Intrinsic motivation towards physical activity was measured with a self-designed parent survey. Children also completed self-designed surveys on their own intrinsic motivation for physical activity and perceived social support from peers and parents.	Wearing of the Fit-bits was recorded (10+ hours = "wear day")	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Empowering patients to perform physical therapy at home	Adherence to physical therapy exercises	Self-designed battery including: One item on pain perception, "Rate the pain you feel in your knee now"; One item on interactions, "The interactions with the environment felt natural" ; Two items on attention demand = "I could concentrate on the assigned exercise", "I had to think about the mechanisms (sensors and tablet) when performing the exercises"; Two items on learning, "The program enabled me to improve my performance", "For the unfamiliar exercise, this program helped me learn"; Four items on control, "I was able to control the movements of the avatar's leg in the game", "There was a demonstrator showing me the exercises", "The avatar that represented me was on the left of the screen", "I prefer not having a demonstrator in the scene"; Three items on usability,"I can perform the entire program on my own at home without help", "I prefer this system over written home exercise instruction", "Overall, I am completely satisfied with this experience"; Two items on time perception, "How long do you think the session took?", "During the session, did you ever feel you lost track of time?". One item on usability, "Do you think this system would make you want to do the exercises?"."	Average performance of the exercises	Motion collected in real time using a three axis positioning miniature wireless accelerometer.
4	Virtual Self-Conversation using Motivational Interviewing Techniques to Promote Healthy Eating and Physical Activity: A Usability Study	Promotion of healthy eating and physical activity in People Living With Obesity (PLWO)	Readiness Rulers (RR); The Suitability Evaluation Questionnaire (SEQ); Subjective rating of the illusion of body ownership through a 7-point Likert scale.	N/A	N/A

J2. Encouraging Physical Activity: Perceived Exertion and Motivation

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Acute Effect of Virtual Reality Exercise Bike Games on College Student's Physiological and Psychological outcomes	Perceived exertion, self-efficacy, and enjoyment	Two items assessing previous experience with VR exercise games and, if applicable, self-perceived performance; Self-rating using the Borg Rating of Perceived Exertion (RPE) Scale assessing subjective exercise intensity every four minutes during each 20-minute exercise session; Self-designed survey assessing self-efficacy after each exercise session through rating of three statements (e.g., "...I have confidence in my performance in this activity"); Self-designed survey assessing enjoyment after each exercise session through rating of five statements (e.g., "I wish I could play this activity more than I get the chance to").	N/A	Blood pressure

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
2	Virtual-Reality Exergaming Improves Performance During High Intensity Interval Training	Perceived exertion, intrinsic motivation, subjective vitality, and exercise intentions	Rating of perceived exertion (CR-10); Intrinsic Motivation Inventory (IMI), specifically the subscales of Interest/Enjoyment, Effort/Importance, and Competence; Subjective Vitality Scale, 5-item version; Intentions to engage in the exercise just completed over the next month were assessed using a 2-item measure, (1) "I intend to engage in the type of exercise I performed today at least 3 times per week during the next month", (2) "I intend to engage in the type of exercise I performed today at least 5 times per week during the next month".	Mean power output during high-intensity interval training (HIIT)	Measurement of expired gases and continuous measurement of VO2 during exercise

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	A Second Chance at Health: How a 3D Virtual World Can Improve Health Self-Efficacy for Weight Loss Management Among Adults	Health self-efficacy	Abbreviated version of the Self-Efficacy and Exercise Habits Survey; 5-item 5-a-day Fruits and Vegetables Confidence Scale from the abbreviated version of the Self-Efficacy and Eating Habits Scale. Self-reported frequency of exercise and self-reported 5-a-day eating behavior (number of servings of fruits and vegetables they ate on average/day). Self-reports on self-presence and perceived effects of their avatar on health self-efficacy (i.e., avatar efficacy). Three open-ended questions: "Do you think Second Life (SL), the social virtual world, has affected your nutrition, physical activity, and/or view of your body? Please explain below." (2) "What challenges, if any, do you face when using SL?" (3) "Please sum up your experience using SL for this study. How did using SL affect your health and well-being?"	Self-reported frequency of exercise and self-reported 5-a-day eating behavior (number of servings of fruits and vegetables they ate on average/day).	Weight, height, and BMI.

J3. Psychological Aspects of Eating and Body Image: Body Image Attitudes and Behaviours

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Developing a novel measure of body satisfaction using virtual reality	Body image disturbance (BID)	State-level Body Image Dissatisfaction (state BID); Weight Concerns Scale (WCS); Body Mass Index (BMI); Eating Disorder Examination Questionnaire (EDE-Q).	Minimum interpersonal distance, visual attention, and approach preference towards avatars of each size	N/A
2	Virtual Reality Environment for Body Image Modification: A Multidimensional Therapy for the Treatment of Body Image in Obesity and Related Pathologies	Body image attitudes and behaviour in obesity	Italian version of the Body Satisfaction Scale (BSS); Italian version of the Body Image Avoidance Questionnaire (BIAQ); The Figure Rating Scale (FRS), and The Contour Drawing Rating Scale (CDRS).	Semi-structured interview based on the Italian version of the Eating Disorders Examination for the clinical interview.	N/A
3	A Virtual Reality Full Body Illusion Improves Body Image Disturbance in Anorexia Nervosa	Body size estimation of body parts	Body Attitude Test (BAT); Eating Disorder Inventory-II (EDI-II); Embodiment Questionnaire (EQ).	Participants were asked to estimate "the width and circumference of several parts of their body (pre size estimation). Width of the shoulders, abdomen and hips was estimated by placing two adhesive markers on the wall representing the left and right side of the body. Circumference was estimated using a piece of string that was placed on the floor so that it would fit exactly around each body part (...) participants estimated their height using an adhesive marker they placed on the wall."	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Effectiveness of Cognitive Behavioral Therapy Supported by Virtual Reality in the Treatment of Body Image in Eating Disorders: One year follow-up	Body image and eating disorder psychopathology	Body image measures included: Body Image Automatic Thoughts Questionnaire (BIATQ); Body Areas Satisfaction Scale (BASS); Situational Inventory of Body-Image Dysphoria (SIBID). Eating disorder psychopathology measures included: The Bulimic Investigatory Test-Edinburgh (BITE); The Eating Attitudes Test (EAT).	Body image measures included: SCID I Interview; SCID-II Interview; Body Attitude Test (BAT). Weekly evaluation of symptoms, including frequency of binges and purgative behavior, "degree of fear and avoidance of forbidden food, as well as the degree of discomfort with one's body in front of a mirror".	N/A
5	Using Immersive Virtual Reality to Modify Body Image	Perceptual boundary between what participants classify as a fat versus a thin body	The Eating Disorders Examination Questionnaire (EDE-Q); 16-item Body Shape Questionnaire (BSQ-16b); Beck Depression Inventory (BDI); Rosenberg Self-Esteem Scale (RSE).	N/A	Body mass index (BMI)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
6	Resize Me! Exploring the User Experience of Embodied Realistic Modulatable Avatars for Body Image Intervention in Virtual Reality	Body image (including body awareness and body weight perception)	Body scan experience was assessed through a self-designed semi-structured interview, specifically concerning expectations, physical and psychological dis-/comfort and assessment of their body measures*; Self-designed questionnaire assessing VR experience and its overall UX*; Self-designed semi-structured interview assessing (1) participant expectations and feelings toward the avatar, (2) favoured body weight modification method and perceived difficulty of the body weight estimation, (3) intensity of body awareness, and (4) affect towards their body. Igroup Presence Questionnaire (IPQ)*; Two items assessing embodiment, specifically VBO and agency, based on items of the Virtual Embodiment Questionnaire (VEQ); Simulator Sickness Questionnaire (SSQ)*; Revised version of the Uncanny Valley Index (UVI); One item assessing workload on the SEA scale, a German version of the Rating Scale Mental Effort. One item assessing body awareness, derived from the State of Mindfulness Scale (SMS).	Participants ordered three body weight modification methods concerning their workload, perceived body weight estimation difficulty, vividness, contentment, and overall preference; Efficiency of avatar calibration was assessed through the average time spanning from the beginning of calibration until the end; Passive Body Weight Estimation (PET) was used to assess the "participant's ability to estimate body weight based on visual perception of avatars with modified body shapes"; Active body weight estimation (AMT) was used to assess the "participant's ability to modify the avatar's body weight to match a requested numeric value".	N/A

J4. Psychological Aspects of Eating and Body Image: Eating Behaviour and Perception

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	The Appeal of Virtual Chocolate: A Systematic Comparison of Psychological and Physiological Food Cue Responses to Virtual and Real Food	Food cues responses (FCR)	Craving (on a scale from 0 to 100).	N/A	Salivation (volume measured in grams)
2	A Randomized Trial Testing the Effectiveness of Virtual Reality as a Tool for Pro-Environmental Dietary Change	Change in dietary carbon footprint and self-efficacy	Food frequency questionnaire; Self-designed tool with two items assessing self-efficacy adapted from Huang et al., two items on eating behaviour, three items on response efficacy, four items on intentions, and five items on psychological distance.	N/A	N/A

J5. Reducing Medical Symptom Discomfort

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Effect of Virtual Reality on Sleep-Deprived Individuals	Sleep deprivation symptoms	PSQI scale	N/A	N/A
2	A Clinical Trial of a Patient-Customized Virtual Reality Intervention for Tinnitus	Tinnitus-related distress, quality of life, and sleep quality	Tinnitus Handicap Inventory (THI); Pittsburg Sleep Quality Index (PSQI); World Health Organisation Quality of Life Questionnaire (WHO-QoL); Profile of Mood States (POMS); Hospital Anxiety and Depression Scale (HADS); Simulator Sickness Questionnaire (SSQ)*.	N/A	EEG

Appendix K

Social Well-Being: Dependent Variables & Outcome Measures

K1. Enhancing Facial Emotion Processing: Impact of Virtual Abuse on Emotional Processing

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Being the Victim of Virtual Abuse Changes Default Mode Network Responses to Emotional Expressions	Processing and responding towards emotional faces	Interpersonal Reactivity Index (IRI); Buss-Perry Aggressiveness Questionnaire (BPAQ).	N/A	Changes in Default Mode Network (DMN) activity, fMRI, anatomical images and a separate functional localizer

K2. Enhancing Facial Emotion Processing: Emotion Recognition in Clinical Populations

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
2	Design of a Virtual Reality System for Affect Analysis in Facial Expressions(VR-SAAFE); Application to Schizophrenia	Processing and responding towards emotional faces	Screening measures included: North American Adult Reading Test (NAART) and IQ. Ourcome measures included: Brief Psychiatric Rating Scale (BPRS); Scale for the Assessment of Positive Symptoms (SAPS); Scale for the Assessment of Negative Symptoms (SANS); Schizotypal Personality Questionnaire (SPQ); Chlorpromazine (CPZ) Equivalent Dose.	Semi-structured clinical interviews assessing symptoms over the past month.	N/A
3	Assessing the Utility of a Virtual Environment for Enhancing Facial Affect Recognition in Adolescents with Autism	Facial affect recognition	Social Responsiveness Scale (SRS); Social Communication Questionnaire—Lifetime Version (SCQ).	The identification of facial expressions of avatars was evaluated through accuracy, response latency, and confidence ratings.	Gaze pattern measures included eye tracking when attempting to identify different kinds of facial expressions in avatars.

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments	Emotion recognition	Autism Diagnostic Observation Schedule-Generic (ADOS-G); Social Responsiveness Scale (SRS) filled by parents; Social Communication Questionnaire (SCQ) filled by parents, Wechsler Abbreviated Scale of Intelligence (WASI).	N/A	Electrocardiogram (ECG), pulse plethesymogram (PPG), skin temperature (SKT), galvanic skin response (GSR), blink rate and pupil diameter, and eye gaze when presented with emotional expression stimuli
5	Brain Responses to Biological Motion Predict Treatment Outcome in Young Adults with Autism Receiving Virtual Reality Social Cognition Training: Preliminary Findings	Emotion recognition and theory of mind (ToM)	Emotion recognition measures included: The Advanced Clinical Solutions for WAIS-IV and WMS-IV Social Perception Subtest (ACS-SP)	Theory of Mind measures included: The Social Attribution Task where accuracy and attribution aspects were scored separately.	BOLD (blood oxygen level dependent) responses using a biological motion fMRI task to measure social information processing.

K3. Enhancing Social Functioning: Social Connectedness

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Familiarity with Teammate's Attitudes Improves Team Performance in Virtual Reality	Joint decision-making	Attitude familiarity was assessed by asking participants to indicate their attitudes toward 27 objects related to the task (e.g., behaviors, situations, and objects)	N/A	N/A
2	Immersive Reality Experience Technology for Reducing Social Isolation and Improving Social Connectedness and Well-being of Children and Young People Who Are Hospitalized: Open Trial	Social inclusion, social connectedness, and well-being	Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS); Social Connectedness Scale; Social Inclusion Scale.	N/A	N/A
3	Synchrony and Social Connection in Immersive Virtual Reality	Self-report social closeness, proximity, and mimicry during idle period	Mini-IPIP Personality scale; Positive and Negative Affect Scale (PANAS); Social closeness was averaged from self-ratings on connectedness, likeability, and ratings of similarity in personality from the adapted version of the Inclusion of Other in Self (IOS) Scale.	Physical proximity (in meters) to both virtual characters at each frame of the VR session was identified; [...] instances of mimicry were also recorded to evaluate whether mimicry occurred more frequently following synchrony compared to non-synchrony.	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
4	Dynamic Interactive Social Cognition Training in Virtual Reality (DiSCoVR) versus Virtual Reality Relaxation (VRelax) for People with a Psychotic Disorder: A Single-Blind Multicenter Randomized Controlled Trial	Social cognition and social functioning	Personal and Social Performance (PSP) scale; Positive and Negative Syndrome Scale (PANSS); The Social Interaction Anxiety Scale (SIAS); The Green et al. Paranoid Thought Scales; Beck Depression Inventory (BDI); Beck Anxiety Inventory (BAI); Perceived Stress Scale (PSS); the Self Esteem Rating Scale (SERS); The Simulator Sickness Questionnaire (SSQ)*.	Ekman 60 Faces Test; The Awareness of Social Inference Task (TASIT); Experience Sampling Method (ESM) diaries where participants received 10 daily text messages at semi-random moments for 7 days; National Adult Reading Test (NART); Trail Making Test (TMT)	N/A
5	Study Protocol for a Randomized Controlled Trial of RealConsent2.0: A Web-based Intervention to Promote Prosocial Alcohol-Involved Bystander Behavior in Young Men	Prosocial bystander behaviour in alcohol-specific context	Bystander Behaviour Scale (BBS) Self-Report Scale; Barriers to Sexual Assault Bystander Intervention Scale; Outcome Expectancies for Intervening Scale; Bystander Efficacy Scale; Bystander Attitude Scale; Sexual Experience Survey Short Form Perpetration; Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5); Daily Drinking Questionnaire; Alcohol Expectancies Regarding Sex, Aggression, and Sexual Vulnerability Questionnaire; Revised Conflict Tactics Scale; Rape Empathy Scale; Differential Reinforcement Scale; Peer Density Network Scale.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
6	A Virtual Model of Migrant Students' Social and Living Skills Development in the Conditions of the Host University: Need Analysis of Migrant Students	Social and living skills and abilities	N/A	Structured interview to identify "barriers that migrant students face within the host educational institution in the receiving country"; Focus-group assessing (1) the 'Social Mirror' simulator architecture and efficiency, and (2) the identification of improvements in the 'Social Integrator' VR simulator.	N/A

K4. Enhancing Social Functioning: Empathy and Perspective-Taking

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
7	Investigating the Influence of Inter-group Contact in Virtual Reality on Empathy: An Exploratory Study Using AltspaceVR	Empathy	Affective Empathy Scale; Body ownership was assessed using items from Peck et al. (2013)'s questionnaire; Four item scale assessing co-presence in VR to measure the degree of salience of the other person in the interaction.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
8	The Enemy's Gaze: Immersive Virtual Environments Enhance Peace Promoting Attitudes and Emotions in Violent Inter-group Conflicts	Perspective-taking in the context of inter-group relations	Battery: Three items assessing empathetic emotions and fear, specifically the degree to which participants experienced emotions toward the out-group members and fear of the Palestinian couple; Four items assessing positive appraisals, specifically the degree to which the Palestinian couple was perceived positively; One item assessing attribution of future benign intentions, specifically the likelihood that the Palestinian man intended to do as he stated when reaching into his jacket: "The Palestinian intended to issue documents"; Two items assessing support for compensation, specifically the degree to which participants supported compensation for the Palestinian couple in case they were misidentified as militants and mistakenly shot and wounded by the soldiers (e.g., "I would support providing financial compensation to the Palestinians for harming them" and "I would support an official Israeli apology to the Palestinians for harming them"); One item assessing political ideology, ranging from 1—extreme right, to 7—extreme left).	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
9	An Auti-Sim Intervention: The Role of Perspective Taking in Combating Public Stigma with Virtual Simulations	Perspective taking, emotional concern, helping intentions, willingness to volunteer, and attitudes toward individuals with ASD	Battery: Six item scale originally modified from Cohen's Identification scale; Four items of a sympathy scale adapted from the Pity sub-scale of the Attribution Questionnaire (AQ); 15 items from the 16 item Societal Attitudes sub-scale of the Societal Attitudes Towards Autism (SATA) scale; Six item modified version of the Helping Behavioural Intention Scale; 10 adapted situations from the The Level of Contact Report; Rating of personal agreement with the statement "I would be willing to volunteer at the Autism Society of America".	N/A	N/A

K5. Enhancing Social Functioning: Social Skills in Autism Spectrum Disorder (ASD)

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
10	Understanding Performance and Verbal-Communication of Children with ASD in a Collaborative Virtual Environment	Social communication	Self-designed questionnaire assessing satisfaction with the game play experience: (Q1) "How much did you like playing the games with your partner?"; (Q2) "Overall, how would you rate the difficulty of the games"; (Q3) "Overall, how do you think you played", (Q4) "Was it easy to talk to your partner in order to figure out how to complete the puzzle?"; (Q5) "Did you get better at talking with your partner the longer you played?"; (Q6) "Did your performance change at all by the end of today's visit? That is, as you played more games, did your performance get better or worse?"	Participant speech were coded offline to assess communication patterns, specifically: frequency of words, frequency of question asking, frequency of information sharing-response, frequency of information sharing-spontaneous, frequency of social reinforcement-positive, frequency of social reinforcement-negative, frequency of directives, frequency of socially oriented utterance, and frequency of speaker changes. Performance analysis on the Collaborative Virtual Environment(CVE) was assessed using four variables: success frequency, time duration, collaborative action duration, and collaborative movement ratio.	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
11	Multi-modal Adaptive Social Interaction in Virtual Environment (MASI-VR) for Children with Autism Spectrum Disorders (ASD)	Emotional expression during social interactions	Social Responsiveness Scale (SRS) completed by parents; Social Communication Questionnaire (SCQ) completed by parents	Autism Diagnostic Observation Schedule-Generic (ADOS-G); Facial recognition measure called NEPSY together with an isolated facial expression recognition measure	Eye gaze monitoring and fixation duration, electrocardiogram (ECG), pulse plethesymogram (PPG), skin temperature (SKT), galvanic skin response (GSR), electromyogram (EMG), respiration (RSP), electroencephalogram (EEG) monitoring, blink rate, and pupil diameter"
12	Design and Development of a Virtual Dolphinarium for Children with Autism	Nonverbal communication skills	Test of Nonverbal Intelligence-Third Edition (TONI-3); Gilliam Autism Rating Scale (GARS).	Six tasks were used to screen for Functional Development Learning and Behavioral Suitability for VDI: (a) identifying parts of a dolphin, (2) identifying shapes traced on palm, (3) geometric shape copying, (4) copy-drawing of a dolphin, (5) dolphin maze–draw a line to lead the dolphin through the maze to coral reef located in the middle of it, (6) name writing and sentence copying.	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
13	Design of a virtual reality based adaptive response technology for children with Autism	Social communication skills	Screening measures included: Adequate language skills; Social Responsive Scale (SRS); Social Communication Questionnaire (SCQ); Autism Diagnostic Interview-Revised (ADI-R).	Autism Diagnostic Observation Scheduled-Generic (ADOS-G)	Gaze data along with task-related event markers (e.g., trial start and trial stop, participant's response, etc.)
14	Design of Gaze Sensitive Virtual Social Interactive System for Children with Autism	Social interaction and eye-contact	Screening measures included: Clinical diagnoses of ASD; Score less (or equal to) than 80 on the Peabody Picture Vocabulary Test-3rd Edition (PPVT-III); Social Communication Questionnaire (SCQ); Social Responsiveness Scale (SRS).	N/A	Gaze data, including pupil diameter (PD), mean blink rate (BR), sum of fixation counts (SFC) and average fixation duration for each region of interest (ROI).

K5. Enhancing Social Functioning: Social Cognition and Functioning in Psychotic Disorders

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
15	A Virtual Reality-Integrated Program for Improving Social Skills in Patients with Schizophrenia: A Pilot Study	Social skills	Positive and Negative Symptoms Scale (PANSS); Assertion Inventory (AI); Social Avoidance and Distress Scale (SADS); Social Functioning Scale (SFS).	Simulated Social Interaction Test (SSIT); Committed errors were recorded (errors made in the facial recognition task, e.g., identify the wrong emotion); Assertive behaviours were recorded (correct identification of assertive, passive, and aggressive behaviours/communication styles of others; Time spent in a conversation with the avatar was recorded. Assertive behaviours and time spent in conversations were recorded	N/A
16	Neurocognition, Presence and Acceptance of a VR Programme for Psychotic Patients: A Correlational Study	Social skills	Screen for Cognitive Impairment in Psychiatry (SCIP)-Spanish version; Slater-Usoh-Steed Presence Questionnaire (SUS)*.	Continuous Performance Test (CPT II version 5 for Windows).	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
17	Online Social Cognition Training in Schizophrenia: A Double-Blind, Randomized, Controlled Multi-Site Clinical Trial	Social cognitive abilities and functional capacity	The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT); The Empathetic Accuracy (EA) Task; Positive and Negative Syndrome Scale (PANSS); Global Functioning Scale: Social and Role (GFS); Social Funcitioning Scale (SFS); An abbreviated version of the Quality of Life Scale (QLS); the Specific Levels of Functioning Scale (SLOF); The Ambiguous Intentions Hostility Questionnaire (AIHQ).	The Penn Emotional Recognition Test (ER40); The Prosody Identification Test (PROID); UCSD Performance-based Skills Assessment (UPSA-2); The Penn Faces Memory Test (PFMT); The Morphed Faces Task; The Awareness of Social Inference Test (TASIT); The Faux Pas Recognition Test; The Source Memory Test; Virtual Reality Functional Capacity Assessment Tool (VRFCAT)	N/A
18	Virtual-Reality-Based Social Cognition and Interaction Training for Patients with Schizophrenia: A Preliminary Efficacy Study	Social cognition and social functioning in patients with schizophrenia	Positive and Negative Syndrome Scale (PANSS); Chinese version of the Personal and Social Performance Scale (PSP); Chinese version of Social Cognition Screening Questionnaire (C-SCSQ).	Digit Span Test (DST); Verbal Fluency Test (VFT); Chinese version of the Face-Affective Identification Task (C-FAIT).	N/A
19	The Acceptability and Feasibility of a Novel Virtual Reality based Social Skills Training Game for Schizophrenia: Preliminary Findings	Social attention	Brief Psychiatric Rating Scale (BPRS); Scale for the Assessment of Positive Symptoms (SAPS); Scale for the Assessment of Negative Symptoms (SANS); Social Functioning Scale (SFS).	National Adult Reading Test, Revised (NART-R)	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
20	VR-SOAP, a Modular Virtual Reality Treatment for Improving Social Activities and Participation of Young People with Psychosis: A Study Protocol for a Single-blind Multi-centre Randomized Controlled Trial	Quantity and quality of social contacts, leisure activities and social participation	Secondary outcome measures included: Social Functioning Scale (SFS); The Brief Negative Symptom Scale (BNSS); The Revised Green Persecutory Thoughts Scale; The Social Interaction Anxiety Scale (SIAS); The Self-esteem Rating Scale Short-form; The Internalized Stigma of Mental Illness scale; The Inventory of Interpersonal situations (ISS); The Dating and Assertion Questionnaire; Positive and Negative Syndrome Scale (PANSS); I-group Presence Questionnaire (IPQ)*.	Primary outcome measures included: 35 item ESM diaries where the participant is asked at semi-random moments whether they have engaged in any (social) activity, at that particular moment, and since the previous questionnaire– prompted by a signal on their phone, five times a day for 2 weeks). Secondary outcome measures included: The Hinting Task; The Dutch version of the Bell and Lysaker emotion recognition task (BLERT).	N/A

Appendix L

Behavioural and Attitudinal Change: Dependent Variables & Outcome Measures

L1. Behavioural and Attitudinal Change: Behavioural Change

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
1	Encouraging Bystander Helping Behaviour in a Violent Incident: A Virtual Reality Study Using Reinforcement Learning	Learning	N/A	"Any action, verbal or physical that the participant attempted to get the attention of the avatars", (e.g., reaching out to them, moving close to them, walking in between the avatars, moving into their field of view, waving a hand or any hand gesture directed to the avatars).	N/A
2	Effect of Virtual Reality Aggression Prevention Training for Forensic Psychiatric Patients (VRAPT): Study Protocol of a Multi-center RCT	Aggression	Aggression Questionnaire (AVL); Childhood Trauma Questionnaire – Short Form (CTQ-SF); Novaco Anger Scale and Provocation Inventory (NAS-PI); State-Trait Anger Expression Inventory-2 (STAXI-2); Barratt Impulsiveness Scale (BIS-11); Buss-Durkee Hostility Inventory-Dutch (BDHI-D); Reactive Proactive Questionnaire (RPQ); I-Group Presence Questionnaire (IPQ)*.	Social Dysfunction and Aggression Scale (SDAS)– completed weekly by staff; Hostile Interpretation Bias Task (HIBT)	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
3	Evaluation of a Virtual Reality Enhanced Bullying Prevention Curriculum Pilot Trial	Reduction in bullying behaviour	Teen Conflict Scale; Psychological Sense of School Members Scale; The University of Illinois Willingness to Intervene in Bullying Episodes; Nine-item Illinois Bully Scale; Relational Aggression Perpetration Scale; Cyberbullying perpetration was assessed with a four item scale based on Ybarra, Espelage, and Mitchell (2007).	N/A	N/A
4	The Impact of Vivid Messages on Reducing Energy Consumption Related to Hot Water Use	Energy-saving behaviours, specifically water temperature and total volume of water used when hand washing	Self-designed vividness scale assessing "how vivid the shower feedback appeared to participants"; Self-designed personalisation scale assessing "how personal the shower feedback appeared to participants".	Water temperature used during hand washing was measured before and after intervention; Total volume of water during hand washing was measured before and after intervention.	N/A

L2. Behavioural and Attitudinal Change: Attitudinal Change

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
5	Using Immersive Virtual Reality to Improve the Beliefs and Intentions of Influenza Vaccine Avoidant 18-to-49-year-olds: Considerations, effects, and Lessons Learned	Perception and belief formation	Self-designed questionnaire assessing participant's history with influenza: (Item 1) whether they were ever told by a doctor or healthcare provider they had influenza; (Item 2) whether they had ever been sick with influenza; (Item 3) whether someone their age should receive an annual flu vaccination; (Item 4) whether in the past 12 months, a doctor or healthcare provider had told them they should receive a flu vaccination; and (Item 5) whether they perceived it easy or difficult to obtain a flu vaccination.; Self-designed questionnaire assessing three potential mediating variables: (1) presence, (2) concern about spreading flu to others, and (3) empathy– and seven dependent variables.	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
6	Investigating the Value of Immersive Virtual Reality Tools for Organizational Training: An Applied International Study in the Biotech Industry	Organisational training/learning	Prior knowledge was assessed with seven questions focused on the content of the training material; Post knowledge was designed to evaluate the training outcomes, specifically one question for each of the dedicated learning activities (24 items); Six questions assessing spatial knowledge; 13 questions assessing conceptual knowledge. Affective measures included: Intrinsic Motivation Inventory (IMI); Self-efficacy Scale adapted from Pintrich et al (1991); Personal value measure was adapted from IMI; Perceived learning measure was based on Hiltz (1994)'s scale.	N/A	N/A

L3. Behavioural and Attitudinal Change: Risk Behaviour and Safety

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
7	Can't Simply Roll it Out: Evaluating a Real-World Virtual Reality Intervention to Reduce Driving Under the Influence	Risk-taking	N/A	DUI behaviour 3 months pre-intervention and 3 months post-intervention.	N/A
8	Evaluating an Intervention to Reduce Risky Driving Behaviors: Taking the Fear Out of Virtual Reality	Risky driving behaviour	Driver Behaviour Questionnaire (DBQ); Emotional Arousal Scale.	The Vienna Risk-Taking Test Traffic: Response latency was used as a measure of propensity for risky driving. Time elapsing between the start of the sequence and the participant's decision to abandon [the sequence] assessed risk-taking inclination in critical road traffic situations (e.g., the longer participants wait to press the button in order to abandon the critical situation, the "higher" the risk taking). "	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
9	Virtual Reality for Public Health: A Study on a VR Intervention to Enhance Occupational Injury Prevention	Behavioural intentions for occupational safety	Self-designed questionnaire to assess experience of immersion; Perceived threat was measured by asking participants to evaluate the "severity of tractor rollover accidents (e.g., perceived severity) and their chance of being involved in tractor rollover accidents (e.g., perceived susceptibility); Behavioural intentions were measured by agreeing/disagreeing to (S1)"If I operated a tractor in the real world, I would operate the tractor with caution' and (S2)"If I operated a tractor in the real world, I would install a Rollover Protection Structure unless the tractor already has it".	N/A	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
10	Teaching children to cross streets safely: A randomized, controlled trial	Pedestrian safety	N/A	Pre-training measures of children's pedestrian abilities (in laboratory and in the field) were assessed by instructing the children to cross when they perceived the virtual street environment to be safe. Post-training measures were comparable to the pre-training measures. Specific pedestrian safety measures within the assessments included four outcomes: (1) start delay (temporal lag before initiating crossing), (2) hits/close calls (collisions/near-misses with vehicles in simulated crossings), (3) attention to traffic (looks left and right, controlled for time), and (4) missed opportunities (safe crossing opportunities that were missed).	N/A

No.	Title	Dependent Variable/s	Self-report Measures	Behavioural Measures	Physiological Measures
11	Acquisition of Fire Safety Knowledge and Skills with Virtual Reality Simulation	Emergency skills, specifically fire safety skills	Self-designed 10-item Fire Safety Evaluation test assessing existing and acquisition of knowledge of OR fire safety.	Adapted version of the Perioperative Performance Evaluation Tool was used to assess transfer of knowledge of OR fire safety skills: "(1) explaining the elements of the fire triangle, (2) communicating a fire risk assessment, (3) following emergency procedures for a fire, (4) demonstrating PASS (pull-aim-squeeze-sweep) technique, and (5) describing RACE (rescue-alarm-confine-extinguish/evacuate)"; "OR fire drill simulation scenario was video recorded for review and evaluation".	N/A

Appendix M

Overview of Main Categories, Sub-Categories, and Further Classifications

1. Mental Health

1a. Enhancing Psychological Well-being

- Stress and Coping
- Mood States and Emotional Regulation
- Quality of Life and Functionality
- Self-Perception and Agency
- Psychological Well-being in Older Adults
- Mental Health in Extreme Environments

1b. Anxiety and Depression

- Anxiety
- Depression

1c. Symptom Management in Mental Disorders

- Psychotic Disorder Symptoms
- PTSD Symptoms
- Autism Spectrum Disorder (ASD) Symptoms
- Substance Use Disorders

2. Exposure Therapy

2a. Fear of specific phobias

- Acrophobia
- Arachnophobia
- Aviophobia
- Vehophobia
- Nyctophobia
- Claustrophobia

2b. Fear of Medical Procedures

- Dentophobia
- Pre-operative Anxiety

2c. Other phobias

- Public Speaking Anxiety
- Fear of Contamination and Disgust

- Combat-related PTSD

3. Mental Health and Cognitive Functioning

3a. General Population

3b. Elderly Population

4. Cognitive Functioning

4a. Enhancing Cognitive Performance and Attention

4b. Cognitive Health in Aging

4c. Cognitive Performance in Clinical Populations

- Attention--Deficit/Hyperactivity Disorder (ADHD)
- Autism Spectrum Disorder (ASD)
- Schizophrenia
- Post-Traumatic Stress Disorder (PTSD)

5. Health

5a. Encouraging physical activity

- Physical Activity and Exercise Adherence
- Perceived Exertion and Motivation

5b. Psychological Aspects of Eating and Body Image

- Body Image Attitudes and Behaviours
- Eating Behaviour and Perception

5c. Reducing Medical Symptom Discomfort

6. Social Well-being

6a. Enhancing facial emotion processing

- Impact of Virtual Abuse on Emotional Processing
- Emotion Recognition in Clinical Populations

6b. Enhancing social functioning

- Social Connectedness
- Empathy and Perspective-taking
- Social skills in Autism Spectrum Disorder (ASD)

- Social Cognition and Functioning in Psychotic Disorders

7. Behavioural and Attitudinal Change

- Behavioural Change
- Attitudinal Change

- Risk-Behaviour and Safety

Appendix N

Intervention-Related Questionnaires

Immersion

1. Augmented Reality Immersion (ARI) questionnaire: Measures the participant's self-reported immersive experience with the Augmented Reality material.
2. Experience in Immersive Virtual Environment Questionnaire: Assesses the participant's experience in immersive VR: presence, involvement, immersion, flow, usability, skill, emotions, experience effects, judgment, and adopting technology.

Satisfaction

1. Client Satisfaction Questionnaire (CSQ): Assesses the participants' satisfaction with the with the services provided/ intervention they received.
2. The Suitability Evaluation Questionnaire (SEQ): Assesses satisfaction, acceptance, and security of use in VR platforms.
3. Treatment satisfaction questionnaire: Assesses satisfaction with the treatment/VR.
4. Value of VR: Assesses perceived qualities of the virtual experience.

Effectiveness

1. Consumer Reports Effectivenes Scale (CRES-4): Assesses whether patients are satisfied with the therapy they have received and if it has been perceived as effective or not.

Cybersickness

1. CyberSickness in Virtual Reality Questionnaire (CSQ-VR): Assesses the side effects of VR and VR symptoms.
2. Simulator Sickness Questionnaire (SSQ): Assesses the severity and symptoms of simulator sickness experienced by individuals during virtual reality or simulator experiences, helping to evaluate their level of discomfort and potential adverse effects.
3. Virtual Reality Sickness Questionnaire (VRSQ): Assesses the severity and symptoms of simulator sickness or cybersickness experienced by individuals during virtual reality experiences.

Usability

1. IBM Computer System Usability Questionnaire (IBM SCUQ): Measure how usable a computer system is.
2. System Usability Scale (SUS): Assess user friendliness.

Presence

1. Igroup Presence Questionnaire (IPQ): Assesses the sense of presence experienced by individuals in virtual environments.
2. Modified Reality Judgement and Presence Questionnaire (MRJPQ): Assesses perceived qualities of the virtual experience.
3. Presence Questionnaire (PQ): Assesses the sense of presence in virtual scenarios.
4. Slater-Usoh-Steed Presence Questionnaire (SUS): Assesses how realistic participants thought the VR environment was.
5. Spatial Presence Experience Scale (SPEC): Perceived sense of being in the virtual space.

6. Short Presence Scale (SPS) (self-designed): Based on the original MEC-SPQ questionnaire, assesses spatial presence (e.g., I felt as if I were taking part in events, not just watching them).

Side Effects

1. Oxford-VR Side Effects Checklist (O-VRSE): Report whether they had any of the 33 experiences listed in the checklist during the provision of VR therapy.

Restoration Environment

1. Restoration Environment Scale (RES): Assesses the restoration of the design scenarios.

Appendix O

Adverse Events

O1. “Automated VR therapy for improving positive self-beliefs and psychological well-being in young patients with psychosis: A proof of concept evaluation of Phoenix VR self-confidence therapy”

"...include but were not limited to: death; suicide attempts; significant self-harm; any violent incident (needing police involvement); bizarre or unusual behaviour with high risk to self or others; formal complaints about therapy; hospital admission; any event which is life threatening; any event which results in persistent or significant disability or incapacity; any event which is otherwise considered medically significant" (Freeman et al., 2023).

O2. “Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial”

“...include but were not limited to: death, suicide attempts, serious violent incidents, admissions to psychiatric hospital, and formal complaints about the intervention" (Freeman et al., 2018).

Appendix P

Definitions of Terms found in Outcome Measures

1. Self-report measures: Participants providing insight into their own cognition, motivation, behaviour, emotion, or physical state allowing for a differentiated assessment of human thought compared to other psychometric methods (National Collaborating Centre for Mental Health (UK), 2011).
2. Behavioural measures: Regard the observation and notation of a specific aspect of a participant's behaviour by an impartial observer, either in a structured laboratory environment or in a natural environment (Price et al., n.d.).
3. Physiological measures: Consider the recording of any physiological process using a variety of instruments across laboratory or natural settings (Gaffey & Wirth, 2014).
4. Chlorpromazine (CPZ) Equivalent Dose: Measure used in psychiatry to standardize and compare the dosages of various antipsychotic medications based on their potency and efficacy relative to 100 mg of oral dose of chlorpromazine (Venkatasubramanian & Danivas, 2013).
5. Sound-Induced Flash illusion (SIFI): Perceptual phenomenon where observers misperceive the presentation of a single flash of light as multiple flashes due to simultaneous presentation of a different number of auditory beeps (Sun et al., 2022).

Appendix Q

Participant-Related Questionnaires

Q1. Self-Report Measures

Mental Health

- Addiction-related
    1. Addiction Memory Intensity Scale (AMIS): Assesses the intensity of addiction-related memories.
    2. Alcohol Expectancies Regarding Sex, Aggression, and Sexual Vulnerability Questionnaire: Assesses alcohol expectancies for aggression, sexual affect, sexual drive, and susceptibility to sexual coercion.
    3. Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): Assesses alcohol use quantity and frequency in addition to heavy episodic drinking.
    4. Brief Questionnaire of Smoking Urges (QSU): Assesses the intensity of urges to smoke cigarettes.
    5. Canadian Problem Gambling Index (CPGI): Assesses the presence of gambling pathology.
    6. Daily Drinking Questionnaire: Assesses quantity of alcohol consumed and perceived consumption among peers during an average week over the past 3 weeks.
    7. Fagerström Test for Nicotine Dependence (FTND): Assesses the level of nicotine dependence in individuals who smoke cigarettes, providing insights into their addiction severity and potential withdrawal symptoms.

8. Gambling inventory: Identifies the types of gambling activities participants engage in (e.g., lotteries), the frequency of their participation, and the amount of money spent on these activities over the past year.
9. Minnesota Nicotine Withdrawal Scale (MNWS): Assesses withdrawal symptoms.
10. Nicotine Dependence Syndrome Scale (NDSS): Assesses a general factor of nicotine dependence and five, theoretically derived subscales.
11. Obsessive Compulsive Drinking Scale (OCDS): Assesses alcohol craving-related obsessive thoughts and compulsions.
12. Obsessive Compulsive Drinking Scale Questionnaire: Assesses subjective craving for alcohol.
13. The Alcohol Use Disorders Identification Test (AUDIT): Screens for harmful or hazardous alcohol use and assesses alcohol consumption and related behaviours.

- ADHD
    1. Attention Deficit Hyperactivity Disorder Rating Scale (ADHD-RS): Assesses inattentive and hyperactive–impulsive symptoms as well as symptom severity over the previous week.
    2. Conners-3: Assesses ADHD symptoms.
    3. Strengths and Weaknesses of ADHD and Normal Behaviour (SWAN): Assesses behavioural attention.
- Autism Spectrum Disorder (ASD)-related
    1. 10 adapted situations from ‘The Level of Contact Report’: Assesses familiarity with Autism Spectrum Disorder (ASD).

2. 16-item Societal Attitudes subscale of the Societal Attitudes Towards Autism (SATA) scale: Assesses positive attitudes.
3. Autism Behaviour Checklist (ABC): Assesses the presence and severity of behavioural characteristics associated with autism spectrum disorder across sensory, social, and communicative domains.
4. Autism Diagnostic Interview-Revised (ADI-R): Used for autism diagnosis and assessing symptom severity.
5. Autism Spectrum Quotient – Japanese Version (AQ-J): Assesses ASD-specific behaviours and symptoms, including social skills, attention switching, attention to detail, imagination and communication.
6. Childhood Autism Rating Scale (CARS): Quantifies the severity of autism spectrum disorder (ASD) symptoms in children.
7. Childhood Autism Rating Scale 2nd Edition—High Functioning (CARS 2-HF): Assesses ASD severity.
8. Gilliam Autism Rating Scale (GARS): Provides the research team the Autism Quotient (AQ) of each participant.
9. Social Responsiveness Scale - Second Edition: Assesses ASD features.
10. Social Responsiveness Scale (SRS): Assesses the severity of social impairments and autistic traits in individuals.

- Anxiety
    1. Anxiety Disorders Interview Schedule (ADIS-IV): Assesses the history of occurrence of any anxiety disorder based on the DSM-IV criteria.

2. Anxiety Sensitivity Index (ASI): Assesses subjective anxiety sensitivity, regarding the belief that experiencing fear/anxiety causes illness, embarrassment, or additional anxiety.
3. Beck Anxiety Inventory (BAI): Assesses anxiety symptoms.
4. Beck Anxiety Inventory for Youth (BAI-Y): Assesses anxiety symptoms.
5. Children's Anxiety Meter-State (CAMS): Assesses children's current state of anxiety.
6. Depression and Anxiety, and Stress Scales (DASS-21): Assesses the severity of depressive, anxiety, and stress symptoms.
7. Generalised Anxiety Directory (GAD-7): Assesses the severity of generalized anxiety disorder symptoms.
8. Geriatric Anxiety Inventory (GAI): Assesses anxiety symptoms in older adults.
9. Hamilton Anxiety Scale (HAMA): Assesses anxiety severity.
10. Hamilton Rating Scale for Anxiety (HRS-A): Assesses anxiety severity.
11. Hospital Anxiety and Depression Scale (HADS): Assesses the severity of anxiety and depression symptoms in individuals receiving medical care.
12. Liebowitz Social Anxiety Scale-Self Report (LSAS-SR): Assesses fear and avoidance in social situations.
13. Mood and Anxiety Symptom Questionnaire (MASQ): Assesses anxious arousal.
14. Paediatric Anxiety Short Form 8a (PASF): Assesses anxiety levels.
15. Six-item Chinese version of the State–Trait Anxiety Inventory (STAI) questionnaire: Assesses state anxiety.
16. Spence Children’s Anxiety Scale (SCAS): Assesses anxiety.

17. Spielberger Anxiety Questionnaire: Assesses anxiety.
18. State-Trait Anxiety Inventory (STAI): Assesses general levels of anxiety in adults.
19. State-Trait Anxiety Inventory (STAI-Y1 and STAI-Y2) Scales: STAI-Y assesses anxiety consisting of a state anxiety measure (STAI-Y1) and a trait anxiety measure (STAI-Y2).
20. State-Trait Anxiety Inventory for Children (STAI-C): Assesses anxiety in children.
21. The Four-Dimensional Symptom Questionnaire (4DSQ): Assesses four common mental health problems: distress, anxiety, depression, and somatization.
22. Visual analogue scale for anxiety (VAS-A): Assesses anxiety.

- Borderline personality disorder-related
    1. Borderline Personality Features Scale for Children (BPFS-C): Assesses borderline personality traits.
- Depression
    1. Beck Depression Inventory (BDI): Assesses general levels of depression.
    2. Beck Depression Inventory II (BDI-II): Assesses general levels of depression.
    3. Beck Hopelessness Scale (BHS): Assesses feelings of hopelessness in the past week.
    4. Depression and Anxiety, and Stress Scales (DASS-21): Assesses the severity of depressive, anxiety, and stress symptoms.
    5. Emotional State Questionnaire (EST-Q2): Self-report tool for depression screening in Estonia.
    6. Geriatric Depression Scale (GDS): Assesses depression symptoms in older adults.

7. Geriatric Depression Scale-15 (GDS-15): Shorter version of the Geriatric Depression Scale (GDS).
8. Hamilton Depression Scale-17 (HAMD-17): Assesses depression severity.
9. Hospital Anxiety and Depression Scale (HADS): Assesses the severity of anxiety and depression symptoms in individuals receiving medical care.
10. Modified depression literacy scale validated in LMICs: Assesses mental health literacy.
11. Self-rating Depression Scale (SDS): Assesses participant's depression mood.
12. The Four-Dimensional Symptom Questionnaire (4DSQ): Assesses four common mental health problems: distress, anxiety, depression, and somatization.
13. The Montgomery−Åsberg Depression Scale (MADRS): To validate the depressive symptoms in the sample.

- Mental illness
    1. Miller-Forensic Assessment of Symptoms Test (M-FAST): Detects malingered mental illness with seven scales: (1) reported versus observed, (2) extreme symptomatology, (3) rare combinations, (4) unusual hallucinations, (5) unusual symptom course, (6) negative image, (7) suggestibility and (8) apparent lack of cooperation or exaggeration.
    2. Mini-International Neuropsychiatric Interview (MINI): Used for diagnosing psychiatric disorders.
    3. Mini-International Neuropsychiatric Interview-Plus for DSM-IV: Assesses for the presence of comorbid psychiatric disorders.

- OCD-related
    1. Yale-Brown Obsessive Compulsive Scale (Y-BOCS): Assesses OCD symptoms.
- Phobias
    1. Acrophobia Questionnaire (AQ): Assesses fear of heights.
    2. AQ-Anxiety: Assesses anxiety related to fear of heights.
    3. Attitude Toward Heights Inventory (ATHI): Assesses fear of heights.
    4. Danger Expectations and Flying Anxiety Scales (DEFAS): Assesses danger expectations regarding flying and anxiety expectations.
    5. Dental Fear Scale (DFS): Assesses dental trait anxiety
    6. Fear of Flying Questionnaire (FFQ): Self-report questionnaire where participants rate their level of fear or discomfort in various flight-related situations
    7. Fear of Spiders Questionnaire (FSQ): Assesses for fear of and beliefs about spiders.
    8. Fear Thermometer questionnaire: Assesses fear of darkness.
    9. Flight Anxiety Modality (FAM): Assesses flight anxiety.
    10. Flight Anxiety Situations (FAS): Assesses flight anxiety.
    11. Heights Interpretation Questionnaire (HIQ): Strongly predicts distress, anxiety, and avoidance of real heights.
    12. Improving Access to Psychological Therapies (IAPT) phobia scale–avoidance: Participants rate their avoidance of heights on a scale from ‘Would not avoid it’ to ‘Always avoid it’.
    13. Mobility Inventory for Agoraphobia –Alone subscale (MIA): Assesses current avoidance for heights.
    14. Modified Dental Anxiety Scale (MDAS): Assesses dental trait anxiety.

15. Oxford Agoraphobic Avoidance Scale (O-AS): Provides an avoidance score by listing “eight simple tasks progressing in difficulty from ‘Stand outside your home on your own for 5 min’ through ‘Travel on your own on the bus for several stops’ to ‘Sit in a café on your own for 10 min’. Participants are asked whether they could do the task now or whether they could not because of anxiety (Yes = 0, No = 1)” (Freeman et al., 2022).
16. Oxford Cognitions and Defences Questionnaire (O-CDQ): Assesses threat cognitions that may contribute to agoraphobia, anxious avoidance, and within-situation defence behaviours.
17. Social Phobia and Anxiety Inventory for Children (SPAI-C): Assesses severity and range of social fears.

- Psychotic-related
    1. Beliefs About Voices Questionnaire-Revised (BAVQ-R): Assesses auditory vocal hallucination levels as well as related beliefs of omnipotence and malevolence.
    2. Brief Psychiatric Rating Scale (BPRS): Assesses the severity of psychiatric symptoms across various domains, such as mood, anxiety, hallucinations, and unusual behaviour.
    3. Dissociative Experiences Scale (DES-II): Measures dissociation, amnesia and absorption.
    4. Negative Voices When Outside: Assesses auditory hallucinations that inhibit a person from participation in everyday social situations.
    5. Paranoia Worries Questionnaire: Assesses the degree to which an individual has been worrying in the past month about others trying to harm them.

6. Paranoia, Obsessive-Compulsive and Depression questionnaire (POD): Combination of three questionnaires, (1) the frequency items from the Paranoia Checklist (PCL), (2) an abbreviated version of the Obsessive Compulsive Inventory – R (OCI-R), and (3) the 20-item Centre for Epidemiologic Studies-Depression Scale.
7. Paranoid Thoughts Scale Part B (GPTS-B): Assesses persecutory thinking over the past month.
8. Psychotic Symptoms Rating Scale (PSYRATS): Assesses auditory vocal hallucination and related beliefs of omnipotence and malevolence.
9. Psychotic Symptoms Rating Scales (PSYRATS-AH) + PSYRATS-AH-Frequency + PSYRATS-AH-Distress: Assesses level of auditory hallucinations, frequency, duration and distress.
10. Revised Green et al. Paranoid Thoughts Scale (R-GPTS): Assesses perceived persecution.
11. Scale for the Assessment of Negative Symptoms (SANS): Assesses the severity and presence of negative symptoms in individuals with schizophrenia and other psychotic disorders, including affective flattening, alogia, avolition-apathy, and anhedonia-asociality.
12. Scale for the Assessment of Positive Symptoms (SAPS): Assesses the severity and presence of positive symptoms in individuals with psychotic disorders, including hallucinations, delusions, and thought disturbances.
13. Schizotypal Personality Questionnaire (SQP): Assesses schizotypal personality traits and features, including unusual beliefs or experiences, social anxiety,

eccentric behaviour, and perceptual distortions, providing insights into personality characteristics associated with schizotypal personality disorder.

14. Screen for Cognitive Impairment in Psychiatry (SCIP)-Spanish version: Assesses cognitive deficits in various psychotic and affective disorders using a battery of tests, including the Working Memory Test (WMT), the Verbal Learning Test-Immediate (VLT-I), the Verbal Fluency Test (VFT), the Verbal Learning Test-Delayed (VLT-II), and the Processing Speed Test (PST).
15. State Social Paranoia Scale (SSPS): Assesses paranoid thoughts about the avatars in VR.
16. The Green et al. Paranoid Thought Scales: Assesses ideas of social reference and ideas of persecution.
17. The Revised Green Persecutory Thoughts Scale: Assesses ideas of reference and 10 items measuring ideas of prosecution.
18. The Revised Mystical Experience Questionnaire (MEQ30): Assesses the intensity of mystical experience.
19. Voice Power Differential Scale: Measure of perceived voice power.
20. Voices Acceptance and Action Scale the VAAS-Action subscale: Assesses of how the patient reacts to his/her voices.

- PTSD

1. Checklist-Military version (PLC-M): Assesses the severity of post-traumatic stress disorder (PTSD) symptoms in military personnel.

2. Posttraumatic Stress Symptom Scale, Self-Report (PSS-SR): Assesses the frequency of PTSD symptoms on three subscales, (1) reexperiencing, (2) avoidance, and (3) arousal symptoms.
3. PTSD Checklist −Military Version (PCL-M): Assesses symptoms of post-traumatic stress disorder (PTSD) specifically in military personnel and veterans.
4. PTSD Checklist for the DSM-5 (PCL-5): Assesses self-reported PTSD symptom severity during the last month.
5. PTSD Symptom Scale (PSS): 17-item self-report scale used to assess PTSD symptoms.
6. The Military Injury Symptom Scale--Military Short Form (MISS-M-SF): Assesses moral injury symptoms that can be used to screen for MI and monitor response to treatment in veterans and active-duty military with or without diagnosed PTSD.

- Trauma-related
    1. Childhood Trauma Questionnaire – Short Form (CTQ-SF): Measures of childhood neglect and abuse.
    2. Life Events Checklist for DSM-5: Identifies exposure to traumatic events.
    3. Peritraumatic Dissociative Experiences Questionnaire (PDEQ): Assesses if a dissociative episode occurred during and/or after the traumatic event that led to PTSD.
    4. The Dissociative Experiences Questionnaire (PDE-Q): Assesses the extent of dissociation at the time of the traumatic event and in the minutes and hours that followed.

5. Trauma-Related Guilt Inventory (TRGI): Assesses 3 domains of trauma related guilt cognitions: (1) global guilt, (2) distress, and (3) guilt cognitions.

Psychological and Emotional Well-being

- Emotional functioning
    1. Cognitive Emotion Regulation Questionnaire (CERQ): Assesses cognitive emotion regulation strategies
    2. Difficulties in Emotion Regulation Scale-18 (DERS-18): Assesses emotional regulation.
        a. Lack of Emotional Awareness subscale: Assesses emotional awareness
        b. Limited Access to Emotion Regulation Strategies subscale: Assesses perceived abilities to manage negative emotions.
    3. Emotional Arousal Scale: Assesses level of emotional arousal.
    4. Emotional Regulation Questionnaire (ERQ): Measures respondents' tendency to regulate their emotions in two ways: cognitive reappraisal and expressive suppression.
    5. Emotional Self-Efficacy Scale: Assesses adolescents' perceived abilities to handle emotional distress.
    6. Implicit Theories of Emotion Scale: Assesses participant's beliefs about emotions.
    7. Japanese version of the Profile of Mood States second edition (POMS2): Assesses seven mood states (anger-hostility, confusion-bewilderment, depression-ejection, fatigue-inertia, tension-anxiety, vigour-activity, and friendliness).
    8. Perth Emotion Regulation Competency Inventory: Assesses emotion regulation.

9. Pictorial Self-Assessment Mannequin (SAM) questionnaire: Assesses emotional states and feelings using visual representations of affective dimensions.
10. Positive and negative affect (PANAS): Assesses positive and negative affect.
11. Positive and Negative Affect Scheduled-Revised (PANAS-X): Assesses negative affect.
12. Positive and Negative Symptoms Scale (PANSS): Assesses the severity and presence of positive and negative symptoms, as well as general psychopathology, in individuals with schizophrenia and related psychotic disorders.
13. Profile of Mood States (POMS): Assesses various mood states, such as tension, depression, anger, vigour, fatigue, and confusion, providing insights into an individual's emotional well-being and psychological state at a given time.
14. The Advanced Clinical Solutions for WAIS-IV and WMS-IV Social Perception Subtest (ACS-SP): Assesses emotion recognition.
15. The Brief Negative Symptom Scale (BNSS): Assesses negative symptoms, examining affect, alogia, asociality, anhedonia and avolition.
16. The Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT): Assesses emotion management.
17. The Self-Assessment Manikin (SAM): Assesses emotional responses in valence (positive/negative), arousal (calm/excited), and dominance (controlled/controlling) areas.
18. Two-Dimensional Mood Scale (TDMS): Assesses mood state (arousal, pleasure, vitality, and stability).

- Fear
    1. Brief Fear of Negative Evaluation Revised Scale (BFNE): Assesses participants' fear of negative evaluation.
    2. Child Fear Scale (CFS): Assesses the level and nature of fears experienced by children.
    3. Fear of Negative Evaluation Scale-Brief Form (FNE-B): Assesses subjective fear of being negatively evaluated by others in social situations.
    4. Fear Record (FR): Participants rate their daily levels of fear, avoidance, and intensity of their main catastrophic thought.
- Happiness and vitality
    1. Mini-Version of the Chinese Happiness Inventory (mini-CHI): Assesses happiness.
    2. Oxford Happiness Inventory: Assesses level of happiness.
    3. Subjective Vitality Scale: Assesses subjective vitality, including (Item 1) At this moment, I feel alive and vital, (Item 2) Currently I feel so alive I just want to burst, (Item 3) At this time I have energy and spirit, (Item 4) At this moment I feel alert and awake, and (Item 5) I feel energised right now.
- Overall Well-being and Functioning
    1. Children's Global Assessment Scale (C-GAS): Assesses overall functioning.
    2. Clinical Global Improvement (CGI-I) scale: Assesses overall global improvement.
    3. Clinical Outcomes in Routing Evaluation-Outcome Measure (CORE-OM): Assess general psychological distress, subjective well-being (symptoms or problems), life functioning, and risk.

4. CORE-SFB: The short version of the Clinical Outcomes in Routing Evaluation-Outcome Measure (CORE-OM).
5. Eurohis Quality of Life Scale (EUROHIS-QOL 8-item index): Assesses subjective quality of life, including overall psychological and behavioural symptomatology.
6. EuroQol 5 dimensions VAS (EQ-5D-3L) Questionnaire: Assesses health related quality of life within five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.
7. Five Facet Mindfulness Questionnaire (FFMQ): Assesses mindfulness on 5 facets, observing, describing, acting with awareness, non-judging of inner experience and non-reactivity to inner experience.
8. Goldberg General health questionnaire (GHQ-12): Detects the presence of psychological distress and psychiatric disorders in the general population by assessing mental well-being, including mood, anxiety, social functioning, and overall psychological health.
9. Mental Health Continuum (MHC-SF): Assesses psychological, emotional, and social well-being.
10. Perceived Restorativeness Scale (PRS-11): Assesses restorativeness.
11. Revised Riff's Psychological Well-being Scale (RPWBS): Assesses six dimensions of psychological well-being among older adults, including autonomy, environmental mastery, personal development, supportive relationships, purpose in life, and acceptance of oneself.
12. RYFF Well-being Scales: Assesses various dimensions of psychological well-being.

13. Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS): Assesses mental well-being.
14. Short well-being scale (self-designed): Composed with selected items from the Mental Well-being Scale and assesses overall mental well-being, including emotional, psychological, and social well-being (e.g. "I am optimistic about the future", I am calm").
15. Shortened version of the Singapore Mental Wellbeing Scale (SMWEB-S): Assesses emotional responses.
16. The Comprehensive Assessment of ACT Processes (CompACT): Assesses psychological flexibility through the subscales openness to experiences (CompACT-OE), behavioural awareness (CompACT-BA), and valued action (CompACT-VA).
17. The Lawton-Brody Instrumental Activities of Daily Living (IADL): Assesses eight functional areas, including utilizing the phone, preparing meals, shopping, cleaning the house, doing laundry, taking medication, using transportation and managing finance.
18. The Personal and Social Performance Scale (PSP): Assessing patients' social and daily functioning.
19. The Specific Levels of Functioning Scale (SLOF): Assesses functioning.
20. Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): Assesses different aspects of positive mental health and overall well-being.
21. WHO Disability Assessment Schedule: Assesses level of functioning.

22. World Health Organization Five Well-being Index (WHO-5): Assesses mental well-being.

- Quality of Life
    1. Abbreviated version of the Quality of Life Scale (QLS): Assesses daily functioning.
    2. Assessment of Quality of Life: Assesses quality of life (QoL).
    3. Health Promotion Administration of the Ministry of Health and Welfare: Assesses life satisfaction.
    4. Lancashire Quality of Life Profile (LQoLP): Assesses participant's subjective life satisfaction.
    5. Life Satisfaction Scale: Participant rates overall degree of satisfaction with life at the present moment (from 1-10).
    6. Maladjustment Scales (MS): Assesses the global impairment that the problem causes in several areas of the participant's life using an 11-point scale (0: "None"; 10: "Extreme").
    7. Maslach Burnout Inventory-Student Survey (MBI-SS): Assesses burnout symptoms and changes the references to work for references to study.
    8. Problem-Related Impairment Questionnaire: Assesses the impairment caused by the disorder on several areas of the participant's life and the degree of change obtained in this aspect after the VR exposure.
    9. Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form: Assesses life satisfaction.

10. Quality of Life Inventory (QoLI): Assesses patients' perception of their quality of life across 15 different domains, including dimensions of interpersonal relationships and social activities.
11. Satisfaction with Life Scale (SWLS): Assesses psychological adjustment.
12. Satisfaction with Life Scale: Assesses quality of life.
13. Short Form Survey (SF-35): Assesses various aspects of health-related quality of life, providing insights into physical and mental well-being across multiple domains.
14. The Hong Kong version of the EuroQol 5-dimensions instrument with a five-level scale (EQ-5D-5L): Assesses health-related QoL.
15. The Quality of Life in Alzheimer's Disease (QOL-AD): Assesses quality of life.
16. The validated Cantril's Ladder of Life: Assesses life satisfaction.
17. World Health Organisation Quality of Life Questionnaire (WHO-QoL): Assesses an individual's perception of their overall quality of life across physical, psychological, social, and environmental domains.

- Stress
    1. Perceived Stress Scale (PSS), Perceived Stress Scale-10 (PSS-10), and Perceived Stress Scale-14 (PS-14): Assesses stress.
    2. The Four-Dimensional Symptom Questionnaire (4DSQ): Assesses four common mental health problems: distress, anxiety, depression, and somatization.
    3. The Psychological Stress Experiences-Short Form 8a (PSES): Assesses stress.

- Worry
    1. Penn State Worry Questionnaire (PSWQ): Assesses anxious apprehension or "worry".
    2. Rumination Reflection Questionnaire (RRQ): Assesses rumination tendencies.

Interpersonal and Social Functioning

- Belonging
    1. Psychological Sense of School Members Scale: Assesses perceived belonging at school.
    2. The Chinese version of the 6-item De Jong Gierveld Loneliness Scale (DJGLS): Assesses loneliness.
    3. The Revised UCLA Three-Item Loneliness Scale: Assesses loneliness.
- Bullying
    1. Nine-item Illinois Bully Scale: Assesses the frequency of traditional bullying perpetration in middle school.
    2. The University of Illinois Willingness to Intervene in Bullying Episodes: Assess student's willingness to intervene when others are being bullied (Ingram et al., 2019).
- Bystander Behaviour
    1. Barriers to Sexual Assault Bystander Intervention Scale: Assesses the participants' self-perceived barriers to intervene while witnessing sexual assault.
    2. Bystander Attitude Scale: Assesses intentions of intervening as a bystander.

3. Bystander Behaviour Scale (BBS) Self-Report Scale: Indicates self-reported bystander behaviour during the past 6 months.
4. Bystander Efficacy Scale: Assesses bystander efficacy.
5. Differential Reinforcement Scale: Assesses peer disapproval for sexual aggression.
6. Outcome Expectancies for Intervening Scale: Assesses positive expectancies for intervening.
7. The University of Illinois Willingness to Intervene in Bullying Episodes: Assess student's willingness to intervene when others are being bullied (Ingram et al., 2019).

- Family functioning
    1. Family Assessment Device: Assesses family functioning.
- Mood States and Affect
    1. Abbreviated Profile of Mood States Questionnaire (POMS): Assesses perceived mood and psychological distress.
    2. Affect Dysregulation Scale: Affect Dysregulation Scale: Assesses the frequency of self-reported difficulties with affect regulation.
    3. Apathy Evaluation Scale (AES) Clinician Version: Assesses objective apathy.
    4. Apathy Evaluation Scale (AES) Self-Rated Version: Assesses subjective apathy.
    5. Feeling and Affect Table to Score Arousal and Appraisal (FAtAAL): Assesses an individual's arousal levels and emotional appraisals in response to stimuli or situations.
    6. Mood and Anxiety Symptom Questionnaire (MASQ): Assesses anxious arousal.

- Public speaking/Communication
    1. Personal Report of Confidence as a Speaker (PRCS-SF): Assesses behavioural responses and affective responses to public-speaking situations.
    2. Personal Report of Communication Apprehension (PRCA-24): Assesses anxiety and fear associated with communicating with other people.
    3. Personal Report of Confidence as a Speaker (PRCS-SF): Assesses behavioural responses and affective responses to public-speaking situations.
    4. Personal Report of Confidence as Public Speaker (PRCS): Assesses degree of public speaking anxiety.
    5. Public Speaking Anxiety scale (PSAS): Assesses the manifestation of cognitive, behavioural and physiological responses to public speaking anxiety.
- Social functioning
    1. Chinese version of Social Cognition Screening Questionnaire (C-SCSQ): Assesses social cognition, including theory of mind, metacognition, and hostile attributional bias.
    2. Chinese version of the Personal and Social Performance Scale (PSP): Assesses social functioning.
    3. Clinician Administered Rating Form of Functional Indicators: Clinician assessment for objective indicators of social functioning, including marital status, employment status, residential status, legal involvement, psychiatric hospitalization, and utilization of medical care.
    4. Ekman 60 Faces Test: Assesses social cognition.

5. Global Functioning Scale– Social and Role (GFS): Assesses participant's social and role functioning by measuring their ability to engage in social interactions and fulfil role responsibilities in daily life.
6. Inclusion of Other in Self (IOS) Scale: Assesses degree of perceived social closeness.
7. Interpersonal Support Evaluation List: Assesses perceptions of social support.
8. Multiple Scale of Perceived Social Support (MSPSS): Assesses perceived social support.
9. Outcomes Questionnaire-45 (OQ-45): Assesses social functioning.
10. Peer Density Network Scale: Assesses individuals' social networks by having participants list five male peers from high school and rate the strength of each relationship on a scale from 0 to 100, with network density calculated as the average relationship strength across these peer connections (Salazar et al., 2019).
11. Social Avoidance and Distress Scale (SADS): Assesses an individual's level of avoidance and distress in social situations, providing insights into social anxiety symptoms and the impact of social interactions on psychological well-being.
12. Social Communication Questionnaire (SCQ): Assesses social communication skills and identify potential deficits associated with autism spectrum disorder (ASD) in children and adolescents.
13. Social Comparison Scale (SCS): Assesses how participants compared themselves to other people or to computer avatars in VR.
14. Social Connectedness Scale: Assesses the degree of interpersonal closeness that individuals feel between themselves and other people, both friends and society.

15. Social Functioning Scale (SFS): Assesses an individual's social functioning across various domains, such as social engagement, interpersonal relationships, and daily functioning.
16. Social Inclusion Scale: Assesses social inclusion that has been validated in young adults and contains three subscales for (1) social isolation, (2) relations, and (3) acceptance.
17. Social Interaction Anxiety Scale (SIAS): Assesses anxiety related to initiating and maintaining social interaction.
18. Social Skills Questionnaire: Assesses social skills.
19. Teen Conflict Scale: Assesses adolescents' ability to listen to, care for, and trust others.
20. The Dating and Assertion Questionnaire: Identifies appropriate dating and assertiveness behaviours.
21. The Inventory of Interpersonal Situations (ISS): Assesses social anxiety and avoidance through 50 hypothetical responses to social situations, scored by the participant based on frequency and discomfort.
22. The Outcome Questionnaire 45 (OQ-45): Assesses social functioning.
23. The Personal and Social Performance Scale (PSP): Assessing patients’ social and daily functioning.

Cognitive and Intellectual Assessment

- Cognitive Examinations
    1. Addenbrooke's Cognitive Examination III (ACE-III): Assesses attention, fluency/language, verbal memory and visuospatial function.

2. Frontal Assessment Battery (FAB): Assesses executive functions– validated for the Portuguese population.
3. Hopkins Verbal Learning Test-- Revised (HVLT-R): Assesses auditory-verbal recall ability.
4. Loewenstein Occupational Therapy Cognitive Assessment-Geriatric (LOTCA-G): Assesses cognitive function, including orientation, visual-spacial perception, visuomotor organization, thinking operation, and attention/concentration functions.
5. Mini-Mental State Examination (MMSE): Assesses global cognitive function.
6. Montreal Cognitive Assessment (MoCA): Assesses cognitive functioning, including global and domain-specific.
7. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Assesses memory, visuospatial skills, language skills, and attention.
8. The Behaviour Rating Inventory of Executive Function-- Adult Version (BRIEF-A): Assesses an individual's perception of their executive functions in nine areas and provides an overall summary score.
9. Neurobehavioural Symptom Inventory (NSI): Assesses overall neurobehavioural symptoms by evaluating self-reported vestibular, somatic, cognitive, and affective post-concussive symptoms (e.g., headache, balance, nausea, etc.).

- Intelligence
    1. Kaufman Brief Intelligence Test - Second Edition: Indicates the Full Scale Intelligence Quotient (FSIQ) for the participants.
    2. Raven's Coloured Progressive Matrices (RCPM): Assesses non-verbal fluid intelligence.

3. Raven's Standard Progressive Matrices (RSPM): Assesses the non-verbal fluid intelligence.
4. Test of Nonverbal Intelligence-Third Edition (TONI-3): Determine their respective intellectual levels in terms of Nonverbal Intelligence Quotient (NIQ).
5. Wechsler Adult Intelligence Scale (WAIS): Determines IQ
6. Wechsler Abbreviated Scale of Intelligence (WASI): Assesses cognitive ability and intelligence
7. Wechsler Intelligence Scale for Children - Fourth Edition (WISC-IV): Assesses cognitive ability and intellectual functioning, including verbal comprehension, perceptual reasoning, working memory, and processing speed.
8. Wechsler Test of Adult Reading: Assesses premorbid verbal intelligence.

Self-perception and Identity

- Self-perception
    1. Brief Core Schema Scale (BCSS): Assesses negative beliefs about the self.
    2. Forms of Self-Criticizing/Attacking and Self-Reassuring Scale (FSCRS): Assesses trait self-criticism and self-compassion
    3. Rosenberg Self-Esteem Scale (RSES): Assesses self-esteem by reflection on current feelings.
    4. Self-Compassion Scale (SCS): Assesses self-compassion across three facets: self-kindness, common humanity and mindfulness.
    5. Self-Compassion Scale – Short Form (SCS-SF): Assesses self-compassion.
    6. Self-Compassion and Self-Criticism Scale (SCCS): Participants are presented with five scenarios likely to induce self-critical or self-compassionate reactions (e.g.,

"You just dropped your new phone and damaged it (scratched)") and are instructed to vividly imagine each scenario. Then, they should rate the extent to which they would react towards themselves in a reassuring/soothing/contemptuous/compassionate/critical/harsh manner.

7. VITA-DMF Self-Efficacy Scale: Assesses self-efficacy.

- Body image/perception and food perception
    1. 16-item Body Shape Questionnaire (BSQ-16b): Assesses body size and shape concerns.
    2. Body Areas Satisfaction Scale (BASS): Assesses the participants' degree of satisfaction and dissatisfaction with regard to 10 body areas.
    3. Body Attitude Test (BAT): Assesses body attitudes.
    4. Body Image Automatic Thoughts Questionnaire (BIATQ): Assesses the cognitive component of body image.
    5. Body Image Avoidance Questionnaire (BIAQ) –Italian version: Assesses avoidance of situations that provoke concern about physical appearance (e.g. tight-fitting clothes, social outings, and physical intimacy).
    6. Body ownership (BOw) questionnaire: Assesses the degree to which individuals feel a sense of ownership over a body or body part, particularly in virtual reality (VR) or experimental settings.
    7. Body Satisfaction Scale (BSS): Assesses body dissatisfaction in eating disorders.
    8. Contour Drawing Rating Scale (CDRS): Assesses body image perception by presenting a set of nine male and female figures with precisely graduated increments between adjacent sizes.

9. Eating Disorder Inventory-II (EDI-II): Assesses eating disorder severity.
10. Eating Disorder Examination Questionnaire (EDE-Q): Assesses the range and severity of eating disorder symptoms, including restraint, eating concern, shape concern, and weight concern.
11. Embodiment Questionnaire (EQ): Three subscales assessing (1) ownership over a virtual body (2) feelings of being in the same location as the virtual body, (3) experiencing agency over the virtual body.
12. Food frequency questionnaire: Assesses an individual's dietary habits and patterns over a specific period by having the participant indicate the frequency and quantity of consumption of various food items and beverages.
13. Situational Inventory of Body-Image Dysphoria (SIBID): Assesses body discomfort and dissatisfaction reactions triggered by behaviour or situations.
14. State-level Body Image Dissatisfaction (state BID): Assesses the level of body image dissatisfaction.
15. The Bulimic Investigatory Test-Edinburgh (BITE): Assesses the cognitive and behavioural characteristics of binge eating disorder and BN.
16. The Eating Attitudes Test (EAT): Evaluates attitudes and behaviour associated with anorexia.
17. The Eating Disorders Examination Questionnaire (EDE-Q): Self-report version of the Eating Disorders Examination (EDE) interview measuring (1) restrictive nature of eating; (2) preoccupation with food and social eating; (3) dissatisfaction with body shape; (4) dissatisfaction with body weight.

18. The Figure Rating Scale (FRS): A set of male and female figures that vary in size from underweight to overweight used as rating scale.
19. Weight Concerns Scale (WCS): Assesses individuals' preoccupations and concerns regarding their body weight and shape.

Coping and Adaptive Strategies

- Coping strategies
    1. Kidcope and Self-Compassion Scale for Youth: Assesses adaptive coping strategies.
- Meditation and mindfulness
    1. Shortened version of the 10-item Experiences During Meditation (EOM-DM) scale: Assesses the experience of meditation.
    2. State of Mindfulness Scale (SMS): Assesses current state mindfulness, specifically present-moment awareness, attention regulation, and an openness toward one's experiences.

Medical and Health-related

- Medical
    1. Medical Outcome Study Short Form-36 Health Survey (SF-36): Assesses health status and functioning along eight dimensions covering (1) functional status, (2) well-being; and (3) overall evaluation of health.
    2. Patient Health Questionnaire (PHQ/-9): Assesses and screens the severity of depressive symptoms in individuals.
    3. Paediatric Anaesthesia Emergency Delirium (PAED): Assesses emergence delirium.

4. Physical Activity Readiness Medical Examination (PARmed-X): Screening tool used to assess an individual's readiness to engage in physical activity or exercise safely.
5. Physical Activity Readiness Questionnaire for Everyone (PAR-Q+): Assesses for medical conditions that can limit physical readiness.
6. The Chinese version of the Edmonton Symptom Assessment System (CESAS): Participant rates the intensity of nine common symptoms experienced by patients diagnosed with cancer.
7. The Medical Outcomes Survey SF-36 (MOS SF-36): Measure health status (broadly defined) and to examine limitations in functioning related to physical activity, social activities, functioning in specific areas, and general health.
8. Tinnitus Handicap Inventory (THI): Assesses the impact of tinnitus on an individual's daily life, including the severity of tinnitus-related distress, functional limitations, and emotional distress caused by tinnitus symptoms.

- Pain
    1. Face, Legs, Activity, Cry, and Consolability (FLACC) scale: Assesses pain intensity.
    2. Pain visual analogue: Assesses level of pain.
    3. Revised Faces Pain Scale (FPS-r): Indicates pain intensity in children.
    4. The Parents’ Postoperative Pain Measure (PPPM): Assesses nonverbal indicators of pain.

5. Wong-Baker Faces Pain Rating Scale (WBFPRS): Assesses pain intensity by asking individuals to select a face that corresponds to their level of pain from a series of facial expressions.

- Sexual risk
    1. 10-item Sexual Risk Knowledge quiz: Assesses HIV-related knowledge, including routes of transmission and course of illness.
    2. Efficacy for Sexual Risk Prevention scale: Assesses the extent to which adolescents "could" engage in specific sexual risk preventive behaviours, such as safe sex discussions with partners.
    3. Revised Conflict Tactics Scale: Assesses sexual violence perpetration in dating relationships.
    4. Sexual Experience Survey Short Form Perpetration: Assesses the number of times that they perpetrated sexual violence using verbal coercion, incapacitation, threats of physical force, and physical force.
    5. The Self-Efficacy for Condom Use scale: Assesses condom use, such as "could use a condom when I'm very upset" and "could use a condom when my partner doesn't want to."
- Exercise-related
    1. Borg Rating of Subjective Exercise Intensity (Perceived Exertion): Assesses an individual's perceived level of exertion or intensity during physical activity.
    2. Godin Leisure-time Exercise Questionnaire (GLTEQ) also known as Godin-Shephard Leisure-Time Physical Activity Questionnaire: Measures the frequency

of strenuous, moderate, and mild leisure physical activity performed for periods of 15 min or more over a usual week.

3. Participant Rating of Perceived Exertion (RPE): Assesses perceived exertion.
4. Rating of perceived exertion (CR-10): Assesses perceived exertion.
5. Self-Efficacy and Exercise Habits Survey: Assesses individuals' beliefs in their ability to engage in exercise and their exercise habits.

- Sleep
    1. Pittsburgh Sleep Quality Index (PSQI): Assesses the quality and patterns of sleep over a one-month time interval, encompassing various dimensions such as sleep duration, latency, disturbances, efficiency, and overall sleep quality.

Personality Traits and Constructs

- Agency
    1. Dispositional Hope Scale (DHS): Assesses agency (the participant’s orientation towards their goals)
- Aggression and Anger
    1. Aggression Questionnaire (AVL): Assesses four sub traits of aggression, i.e., physical aggression, verbal aggression, anger and hostility.
    2. Buss-Durkee Hostility Inventory-Dutch (BDHI-D): Assesses two factors of aggression, covert and indirect aggression.
    3. Buss-Perry Aggressiveness Questionnaire (BPAQ): Assesses individual differences in aggressive tendencies across multiple dimensions, including physical aggression, verbal aggression, anger, and hostility.

4. Novaco Anger Scale and Provocation Inventory (NAS-PI): Assesses anger as a problem of psychological functioning and physical health and to assess therapeutic change.
5. Reactive Proactive Questionnaire (RPQ): Assesses self-reported aggression that is not restricted to a short period of time.
6. Relational Aggression Perpetration Scale: Assesses exclusion, rumour spreading, and other activities meant to damage another child's reputation or social relationships across five items.
7. State-Trait Anger Expression Inventory-2 (STAXI-2): Assesses the experience, expression, and control of anger.
8. The Ambiguous Intentions Hostility Questionnaire (AIHQ): Determined attributional styles.

- Assertiveness and Self-confidence
    1. Assertion Inventory (AI): Assesses assertiveness level, including the ability to express thoughts, feelings, and needs in a direct and respectful manner.
    2. The Self Esteem Rating Scale (SERS): Assesses self-esteem.
    3. The Self-esteem Rating Scale Short-form: Assesses positive and negative self-esteem separately.
    4. Visual analogue scale (VAS) of confidence: Confidence “right now”.
- Attitudes and Beliefs
    1. Personality Disorder Belief Questionnaire (PDBQ): Assesses personality disorder related beliefs.

2. Positive Self Scale (OxPos): Positive self-beliefs: mastery, strength, enjoyment, and character.
3. Questionnaire about participant's history with influenza (self-designed): (Item 1) whether they were ever told by a doctor or healthcare provider they had influenza; (Item 2) whether they had ever been sick with influenza; (Item 3) whether someone their age should receive an annual flu vaccination; (Item 4) whether in the past 12 months, a doctor or healthcare provider had told them they should receive a flu vaccination; and (Item 5) whether they perceived it easy or difficult to obtain a flu vaccination.
4. Questionnaire containing measures of three potential mediating variables -- presence, concern about spreading flu to others, and empathy-- and measures of seven dependent variables (self-designed): Assesses seven variables: (1) community immunity beliefs, (2) flu susceptibility, (3) self-flu severity, (4) other-flu severity, (5) flu vaccination confidence, (6) confidence that one's flu vaccination protects others, and (7) flu vaccination intention.

- Ego
    1. Ego-Dissolution Inventory (EDI): Assesses ego dissolution.
- Engagement with Beauty
    1. Engagement with Beauty Scale: Assesses extent to which participants perceive themselves as emotionally and physiologically responding to beauty in the natural world.

- Helping intentions
    1. Modified version of the 6-item Helping Behavioural Intention Scale: Assesses helping intentions.
- Impulsiveness
    1. Barratt Impulsiveness Scale (BIS-11): Assesses impulse control on three subscales– attentional impulsiveness, non-planning impulsiveness, and motor impulsiveness.
    2. Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency (UPPS-P): Assesses impulsivity.
- Locus of Control
    1. Levenson LoC questionnaire: Assesses locus of control with three subscales, (1) internality, (2) powerful others, (3) chance.
- Motivation
    1. Intrinsic Motivation Inventory (IMI): Assesses intrinsic motivation.
- Optimism
    1. Life Orientation Test Revised (LOT-R): Assesses dispositional optimism.
- Personality
    1. Eysenck Personality Questionnaire (EPQ): Assesses personality.
    2. Mini-IPIP Personality scale: Assesses the Big Five personality traits– openness, conscientiousness, extraversion, agreeableness, and neuroticism.
- Perspective-taking
    1. Cohen's Identification scale: Assesses perspective-taking.

- Resilience
    1. Connor Davidson Resilience Scale (CD-RISC 10): Assesses perceived resilience.
    2. Connor-Davidson Resilience Scale (CD-RISC-25): Assesses perceived resilience within 17 domains.
- Sympathy and empathy
    1. ‘Sympathy scale’ adapted from the Pity subscale of the Attribution Questionnaire (AQ): Assesses emotional concern.
    2. Affective Empathy Scale: Assesses subjective levels of situational empathy.
    3. Interpersonal Reactivity Index (IRI): Assesses different facets of empathy, including cognitive and emotional empathy, perspective-taking, and the ability to understand and share the feelings of others.
    4. Perspective-taking subscale of the Interpersonal Reactivity Index (IRI): Assesses person-centeredness and empathy.
    5. Rape Empathy Scale: Assesses rape-perpetrator empathy.
    6. The Empathetic Accuracy (EA) Task: Assesses empathy.

Miscellaneous

- Academic engagement
    1. Utrecht Work Engagement Survey Scale-Students (UWES-S): Assesses academic engagement that includes three subscales: (1) vigour, (2) dedication, and (3) absorption.
- Insights
    1. Psychological Insight Questionnaire (PIQ): Identifies insights in our sample at the end of the first (control) and second days.

2. Session Rating Scale 3.0 (SRS 3.0): Assesses quality of the therapeutic alliance.
3. Treatment Emergent Symptom Scale (TESS): Assesses the presence and severity of side effects or symptoms that emerge during the course of a treatment (Yu et al., 2023).

- Self-reported Behaviour
    1. Child Behaviour Checklist (CBCL): Assesses competence in activities, social and school domains, specific behavioural problems/diagnoses, and emotional problems.
    2. Driver Behaviour Questionnaire (DBQ): Indicated self-reported engagement in risky drinking behaviours, (a) violations, (b) errors, (c) lapses.
    3. Driving Cognitions Questionnaire (DCQ): Assesses three areas of driving-related concerns: panic-related, accident-related, and social concerns.
    4. Marino Interview Assessment Scale (MIAS): Assesses the degree to which the participant utilized certain skills in an interview setting.
    5. Posttraumatic Avoidance Behaviour Questionnaire (PABQ): Identifies avoidance behaviour in daily life.
    6. Readiness Rulers (RR): Visual Analog Scales ranging from 1 to 10 that assess “importance,” “confidence,” and “readiness” to change.
- Stigma of Mental Illness/ Mental Health Services
    1. Brief Version of the Internalized Stigma of Mental Illness scale: Assesses mental health stigma
    2. Inventory of Attitudes Towards Seeking Mental Health Services: Assesses attitudes toward mental health help-seeking.

3. The Internalized Stigma of Mental Illness scale: Assesses stigma with 29 items on five different subscales: (1) alienation, (2) stereotype endorsement, (3) discrimination experience, (4) social withdrawal, (5) stigma resistance.

- Types of Measurement
    1. Colour Analog Scale (CAS): Assesses subjective experiences, such as pain or mood, by having individuals indicate their level of intensity or preference along a coloured visual analogue scale.
    2. Subjective Units of Discomfort Scale (SUDS): Participants rate their anxiety level on an eleven-point scale (0: no anxiety; 10: extreme anxiety).

Q2. Behavioural Measures

Cognitive Abilities and Executive Functions

- Cognitive Profiling Tools
    1. Brief Neuropsychological Cognitive Examination (BNCE): General cognitive profile that can be used for screening, diagnosis or follow-up.
    2. CNS Vital Signs Cognitive Test: Computerized assessment tool designed to evaluate various aspects of cognitive functioning, including memory, attention, executive function, processing speed, and reaction time, to aid in the assessment of neurological and psychiatric conditions (Gualtieri & Johnson, 2006).
    3. Marino Interview Assessment Scale (MIAS): Assesses the degree to which the participant utilized certain skills in an interview setting.
    4. NEPSY 2nd Edition (NEPSY-II): A battery of 32 individually administered subtests and 4 delayed recall subtests which assess neuropsychological functioning across six domains, including attention and executive functioning, language, memory and learning, sensorimotor, social perception, and visuospatial processing. *In this specific study it is used to assess emotion recognition and theory of mind.
    5. Toulouse Pierón test (TP): Assesses "permanent voluntary attention, concentration, resistance to fatigue, and stimulus processing" (Gamito et al., 2021).
- Memory Assessment Tools
    1. Brief Visuospatial Memory Test-- revised (BVMT-R): Assesses visuospatial learning and memory abilities by presenting a complex figure to the participant and subsequently requiring them to attempt to draw it from memory in the correct location after a delay (Tam & Schmitter-Edgecombe, 2013).

2. Digit Span Test (DST): Neuropsychological background test assessing auditory attention and short-term memory by asking participants to repeat a sequence of digits in the same order as presented or in reverse order (Shen et al., 2022).
3. Level of understanding test (LUT): Assesses how well a participant can understand and remember the content provided in two orientation approaches.
4. Rey Complex Figure (RCF): Assesses visuoconstructive abilities and visual memory by asking participants to reproduce a complex figure from memory (Gamito et al., 2020).
5. Rivermead Behavioural Memory Test (RBMT): Assesses everyday memory performance through tasks that mimic real-life memory demands, such as remembering names, faces, and routes (Tsang & Man, 2013).
6. The Penn Faces Memory Test (PFMT): Assesses immediate and delayed memory for faces.
7. The Source Memory Test: Assesses Memory for the source of items.
8. The Symbol Span Subtest of the Wechsler Memory Scale 4th Edition (WMS-IV): Assesses visual working memory using novel stimuli.
9. WAIS-IV Arithmetic and Digit Span: Assesses working memory.

- Executive Functioning Assessment Tools
    1. Colour-Word Stroop Test (CWST): Assesses executive function through the task of identifying the ink colour of rows of “XXX” and compares the participant’s performance to their ability to identify the ink colour of colour words in conditions where the word and the ink colour are incongruent to each other (e.g., the word “blue” printed in blue ink) (Uttl & Graf, 1997).

2. Trail Making Test (TMT) Parts A and B: Information processing and mental set switching, executive function.
3. Wisconsin Card Sorting Test — Computer Version 4 (WCST-CV4): Assesses executive functioning.
4. Wisconsin Card Sorting Test (WCST/-64): Assesses abstract reasoning and perseveration.

- Processing Speed Assessment Tools
    1. D-KEFS Colour-Word Interference Test Condition 1 (Colour Naming): Assesses processing speed
    2. D-KEFS Colour-Word Interference Test Condition 2 (Word Naming): Assesses processing speed
    3. D-KEFS Trail Making Test, Condition 2 (Number Sequencing): Assesses processing speed
    4. D-KEFS Trail Making Test, Condition 3 (Letter Sequencing): Assesses processing speed
    5. DKEFS Design Fluency Condition 1 (Filled Dots): Assesses processing speed.
    6. DKEFS Design Fluency Condition 2 (Empty Dots): Assesses processing speed.
    7. Trail Making Test (TMT) Parts A and B: Information processing and mental set switching, executive function.
- Cognitive Inhibition Assessment Tools
    1. D-KEFS Colour-Word Interference Test Condition 3: Assesses cognitive inhibition
    2. D-KEFS Colour-Word Interference Test Condition 4: Assesses cognitive inhibition-switching.

- Response Inhibition Assessment Tools
    1. Stroop test: Assesses the ability to inhibit automatic responses and demonstrates cognitive flexibility by presenting a task where reaction time is delayed due to conflicting stimuli.
    2. Stroop colour-word test: Evaluates the ability to inhibit automatic responses and cognitive flexibility by asking participants to name the ink colour of words while ignoring the actual word meanings, which can be congruent or incongruent with the ink colour.
- Cognitive Flexibility Assessment Tools
    1. D-KEFS Trail Making Condition 4 (Letter-Number Sequencing): Assesses cognitive flexibility
    2. Design Condition 3 (Switching Dots): Assesses cognitive flexibility.
    3. Trail Making Test Part B: Set shifting.
- Intelligence Assessment Tools
    1. National Adult Reading Test (NART): Measures premorbid intelligence and intellectual ability.
    2. North American Adult Reading Test (NAART): Evaluates premorbid verbal intelligence in adults by evaluating their ability to correctly pronounce a list of irregularly spelled words, providing an indication of their cognitive functioning prior to any neurological or cognitive impairment.
- Verbal Working Memory and Fluency Assessment Tools
    1. Digit Span Subtest of the Wechsler Adults Intelligence Scale 4th Edition (WAIS-IV): Measures verbal working memory by asking participants to repeat a sequence

of digits in the same order as presented or in reverse order (Mahdieh Sasaninezhad et al., 2024).

2. Verbal Fluency Test (VFT): Neuropsychological background test testing for verbal fluency.

- Creativity Assessment Tools
    1. Guilford's Alternative Uses Test (AU test): Prompts individuals to generate as many alternative and unusual uses as possible for a common object within a specific time frame, measuring divergent thinking abilities and creativity.

Emotional and Psychological Factors

- Facial Emotion Recognition Assessment Tools
    1. Chinese version of the Face-Affective Identification Task (C-FAIT): Assesses ability to perceive and identify emotions from facial expressions.
    2. Ekman 60 Faces Test: Assesses social cognition through the recognition and interpretation of facial expressions across various emotional states.
    3. Emotional Appearance Scale for Children (EASC): Assesses children's ability to express emotions through facial expressions, body language, and overall appearance.
    4. Hostile Interpretation Bias Task (HIBT): 10-min computer task to measure hostile interpretation of ambiguous facial expressions of emotions.
    5. The Dutch version of the Bell and Lysaker emotion recognition task (BLERT): Participants view 35 videos depicting actors narrating situations and expressing emotions or remaining neutral, then choose the corresponding emotion from seven options after each video (Meins et al., 2023).

6. The Penn Emotional Recognition Test (ER40): Assesses facial emotion recognition.

- Affective Observation Tools
    1. Observed Emotion Rating Scale (OERS): Trained professional observes and records the duration of expression for five affect types (pleasure; anger; anxiety; sadness; and general alertness) during a specifies time frame.
    2. Experience Sampling Method (ESM) diaries: Participants received 10 daily text messages at semi-random moments for 7 days on current positive affect, negative affect, stress, overall and in company of others. Enjoyment of current and recent activities, and how the company was experienced, and participant's perceived social cognition.
- Anxiety Assessment Tools
    1. Observational Scale of Behavioural Distress (OSBD) scale: Assesses children's stress and anxiety.
    2. Yale Preoperative Anxiety Scale (mYPAS-SF): Assesses perioperative paediatric anxiety instrument with observational measurements of anxiety in four domains: (1) activity, (2) emotional expressivity, (3) state of arousal, and (4) vocalization (Chamberland et al., 2023).
- Post-Traumatic Stress Disorder (PTSD) Assessment Tools
    1. Clinician-Administered PTSD (CAPS-5) scale for DSM5 ; (CAPS-1): Clinician assessment for objective PTSD presence and symptom severity based on DSM criteria through a structured interview.

2. Emory Treatment Resistance Interview for PTSD: Quantifies level of treatment resistance to evidence-based psychological and pharmacological treatments.

Social and Communication Skills

- Theory of Mind (ToM) Assessment Tools
    1. Baron-Cohen Minds in the Eyes Test (RMET): Assesses Theory of Mind (ToM) based on the participant's ability to attribute emotional states to presented visual stimuli of eyes.
    2. Cartoon stories task: Assesses Theory of Mind by asking participants to interpret and explain social interactions and mental states of characters in cartoon scenarios.
    3. Faux pas test: Assesses ToM, ability to recognize and understand social faux pas situations, and the ability to infer intentions and emotional states of another person.
    4. The Awareness of Social Inference Task (TASIT): Assesses social perception and ToM by observing participants' responses to social situations.
    5. The Faux Pas Recognition Test: Assesses ToM by presenting scenarios where a character unintentionally says something inappropriate, and participants identify the faux pas (Stone et al., 1999).
    6. The Social Attribution Task: Assesses theory of mind.
- Social Cognition and Communication Assessment Tools
    1. The Hinting Task: Assesses social cognition by evaluating the participant's ability to infer the true meaning behind indirect speech through ten short social situations, with each narrative and question read aloud by the test leader (Meins et al., 2023).

2. The Prosody Identification Test (PROID): prosody identification, which involves the ability to perceive and interpret emotional and communicative cues through variations in pitch, rhythm, and intonation in speech (Nahum et al., 2020).
3. Simulated Social Interaction Test (SSIT): Controlled laboratory-based assessment used to simulate social interactions and evaluate social skills, communication abilities, and behavioural responses in individuals.

- Language Assessment Tools
    1. Peabody Picture Vocabulary Test-3rd Edition (PPVT-III): Assesses language skills by presenting the participants with a series of pictures and asks them to select the corresponding word from a set of options, assessing their understanding of spoken vocabulary. In this case, used to ensure that language skills were adequate for the intervention (Lahiri et al., 2011).
    2. The Hungarian Metaphor and Irony test: Assesses pragmatic skills by evaluating the ability of the participant to comprehend metaphors and irony in the Hungarian language.
- Observational Assessments by Third-Parties
    1. Behaviour Rating Inventory Executive Function, 2nd edition (BRIEF-2): Caregiver or teacher-report measure assessing executive function of the participant.
    2. Social Dysfunction and Aggression Scale (SDAS): Systematic recording of staff observations on a broad range of aggressive behaviour on a weekly basis, including but not limited to (1) irritability, e.g., difficulty controlling reactions; (2) negativism, e.g., not wanting to cooperate; (3) directed verbal aggressiveness, e.g., insulting people personally.

3. Social Phobia and Anxiety Inventory for Children-Parent Version (SPAIC-PV): The content and format are identical to the SPAI-C but it is completed by the parent.

Neurodevelopmental and Psychological Disorders

- Autism Spectrum Disorder (ASD) Assessment Tools
    1. Autism Diagnostic Observation Schedule-Generic (ADOS-G): Clinician assessment of communication, social interaction, and repetitive behaviours in individuals suspected of having autism spectrum disorder (ASD), used for diagnosis and treatment planning.
    2. Clancy Autism Behaviour Scale (CABS): Involves the observation of behaviours in children to differentiate between mild, moderate and severe autism.
- Clinical Interviews for Psychiatric Disorder Diagnosis
    1. Structured Clinical Interview for DSM-IV (SCID I and II): Assesses for the presence of other Axis I and II disorders.
    2. Structured Clinical Interview for DSM-IV Axis II (SCID-II): Assesses for personality disorders.
    3. Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID-I): Assesses the diagnosis of SAD and potential comorbidity.
- Alcohol-Related Behaviour Assessment Tools
    1. Moeller et al.'s probabilistic image choice task: Assesses the seeking for alcohol-related stimuli (Barillot et al., 2023).
    2. Soleymani et al.'s Visual Research Test for alcohol-related stimuli: Assesses responses to visual cues related to alcohol consumption (Barillot et al., 2023).

3. Standard Implicit Association Test (IAT): “Measures whether alcohol is more strongly associated with the self or others using images referring to alcohol or neutral images as target categories and words referring to the self or others as attribute categories (eg: ‘me’, ‘I’, ‘myself’, ‘mine’ or ‘they’, ‘them’, ‘their’, ‘others’)” (Barillot et al., 2023).
4. Time Line Follow Back (TLFB) interview: Assesses alcohol use, indicating a total number of standard alcohol drinks per week.

Sensory and Motor Skills

- Visual Processing Assessments
    1. D-KEFS Trail Making Test Condition 1: Assesses visual scanning
    2. The Symbol Span Subtest of the Wechsler Memory Scale 4th Edition (WMS-IV): Assesses visual working memory using novel stimuli.
    3. Trail Making Test Part A: Visual search/selective attention.
- Visual-Motor Integration Assessments
    1. Developmental Test of Visual-Motor Integration (VMI): Assesses the participant’s ability to integrate visual and motor skills by completing visual-spatial tasks.

Behavioural Tendencies and Preferences

- Approach and Avoidance Behaviour Assessment Tools
    1. Approach preference toward avatars of each size: Indicates the participant’s initial choice of avatar group to approach in each populated environment tested by analysing the direction of their first movement in relation to the positions of the avatar groups.

2. Behavioural Avoidance Test (BAT): Assesses the degree of avoidance and fear exhibited in response to specific situations linked to phobias through the observation of an individual's approach towards a feared scenario until they reach a limit (American Psychological Association, n.d.).
3. Liebowitz Social Anxiety Scale (LSAS): Clinician assessment of fear and avoidance of social interaction situations through a structured interview.
4. Minimum interpersonal distance: Measures the smallest absolute distance reached between a participant and each target avatar group).

- Attention Assessment Tools
    1. Attention network task: Assesses cognitive attention.
    2. Continuous Performance Test (CPT II version 5 for Windows®): Computer-based attention-vigilance task which provides information about inattentiveness, impulsivity or difficulties maintaining vigilance.
    3. Digit Span Test (DST): Neuropsychological background test assessing auditory attention and short-term memory by asking participants to repeat a sequence of digits in the same order as presented or in reverse order (Shen et al., 2022).
    4. Digit Vigilance Test (DVT): Assesses sustained attention and psychomotor speed during rapid visual tracking and accurate selection of target stimuli (Tsang & Man, 2013).
    5. Modified little bell test (MTCM): Assesses attention processes (De Luca et al., 2021).
    6. Temporal Order Judgements (TOJ): Comparable to the Simultaneity Judgement (SJ) task, however, participants are asked to determine whether they perceived the

visual (pressing key ‘1’) or auditory (pressing key ‘2’) stimulus as appearing first (Basharat et al., 2023).

7. Visual attention: Total percentage of time within each populated virtual environment that the participant spent looking at each avatar group (Nahum et al., 2020).

- Risk-taking Assessment Tools
    1. Balloon Analogue Risk Task –Youth: Computerised decision-making task assessing risk-taking.
    2. Delay Discounting (DD) task: Measures the preference for smaller and immediate rewards over larger and delayed ones.

Functional and Performance Metrics

- Functional Capacity and Skills Assessment Tools
    1. UCSD Performance-based Skills Assessment (UPSA-2): Assesses functional capacity outcomes, including “skills in 5 areas: household chores, communication, finance, transportation, and planning recreational activities” (Nahum et al., 2020).
    2. Virtual Reality Functional Capacity Assessment Tool (VRFCAT): Functional capacity measure where participants can experience a real-life scenario of a shopping trip through VR
    3. Vocational Cognitive Rating Scale (VCRS): Assesses cognitive functioning in the work environment in patients with severe mental illness as observed by a therapist (Tsang & Man, 2013).

- Performance Assessment Tools
    1. Perioperative Performance Evaluation Tool: Assesses the performance of individuals in perioperative settings, including technical skills, teamwork, communication, and adherence to safety protocols.
    2. The Reliable Digit Span (RDS): In this case it was used as a measure of performance validity, “calculated by summing the longest string of digits correctly recalled on forward plus backward trials (both trials correct) of WAIS-IV digit span” (Pennington et al., 2022).

Miscellaneous

1. Assertive Responding to Voices: Assesses the participant’s ability to respond to voices in an assertive way.
2. The Craving Induction Protocol: Craving induced by personalised cues, then participants rate their craving levels on a VAS ranging from 0 to 10.

Q3. Physiological Measures

Eye Movement and Tracking

1. Average fixation duration: How long the average fixation lasted for, which is the time during which the eyes focus on one object (Galley et al., 2015).
2. Binocular eye movement: Refers to the coordinated movement of both eyes to fixate on and track objects in an environment (Uomori & Yolanda, 1994).
3. Blink rate (BR): Measure of how frequently an individual blinks over a given period of time.
4. Eye tracking data: Saccades, number and duration of fixations and velocity of eye movements.
5. Gaze Behaviour: Identifies the patterns and characteristics of how individuals fixate their gaze on a location in the environment, as well as how they shift their gaze between different locations within the environment (Vickers, 1995).
6. Pupil diameter (PD): Measure of the size of the pupil, which can change in response to light levels, cognitive load, or autonomic nervous system activity (Lin et al., 2001).
7. Spontaneous eye blink rate (sEBR): Measure of the frequency of involuntary blinks that occur naturally, may inform on neurological health, cognitive load and attention, working memory activity, dopaminergic activity, and states of fatigue and stress (Ortega et al., 2022).
8. Sum of Fixation Counts (SFC): Total number of times a person's gaze fixates on a particular location or object within a period of time. Used to assess visual attention, cognitive processing, and interest in stimuli (Banire et al., 2020).

Cardiovascular Measures

1. Blood pressure (BP): Measures the force exerted by circulating blood against the walls of the arteries in an individual. Used to measure autonomic nervous system activation (e.g., during stressful situations).
2. Heart rate (HR): Number of beats per minute reflecting cardiovascular and overall health status. Used to measure autonomic nervous system activation (e.g., stress, anxiety, excitement, etc).
3. Heart rate variability (HRV): The variation in time between consecutive heartbeats.
4. Electrocardiogram (ECG): A recording of the electrical signals in the heart (Mayo Clinic, 2022).
5. Photoplethysmographic (PPG) data: Method of measuring changes in blood volume in tissues, typically by shining a light on the skin and measuring the amount of light that is absorbed or reflected.
6. Pulse plethysmogram (PPG): Non-invasive technique that records the blood volume pulse waveform via optical methods to measure sympathetic arousal.
7. VO2: VO_2 max refers to the maximum amount of oxygen your body can intake and utilize during exercise (Cleveland Clinic, 2022).

Body Measurements

1. Body Mass Index (BMI): Numerical measure calculated by dividing an individual's weight in kilograms by the square of their height in meters, commonly used as an indicator of body fatness and overall health status (Centres for Disease Control and prevention, 2022).
2. Height
3. Waist circumference (WC)
4. Waist-to-hip ratio

5. Weight
6. Body/skin temperature (SKT)

Biochemical and Hormonal Measures

1. Carbohydrate deficient transferrin levels (CDT): A modified version of the iron transport protein transferrin, serving as a highly specific biomarker for heavy alcohol intake. Elevated CDT levels persist for approximately 2-3 weeks after moderate alcohol consumption ((De Vos et al., 2019).
2. Cortisol levels: Cortisol is a hormone made by the adrenal glands and helps the body respond to stress, reduce inflammation, regulate blood sugar and the metabolism, and control blood pressure. In this case used primarily as a measure for stress (MedlinePlus, 2022).
3. Gamma-glutamyl transpeptidase (GGT): An enzyme found mainly in the liver, kidney, pancreas, heart, and brain. Its levels are typically assessed to identify liver dysfunction and serve as an indicator of alcohol consumption (Whitfield, 2001).
4. Haemoglobin changes: Haemoglobin is a protein in red blood cells that delivers oxygen around the body (Billett, 1990).
5. Salivary Alpha Amylase (sAA): The most abundant enzyme in saliva, proposed as a highly indicative biomarker for stress (Nater & Rohleder, 2009).
6. ELISA method: The Enzyme-linked immunosorbent assay method.

Neurological Measures

1. Electroencephalogram (EEG): Measures electrical activity in the brain using small, metal discs (electrodes) attached to the scalp.

2. Electromyography (EMG): Measures muscle response or electrical activity in response to a nerve's stimulation of the muscle.

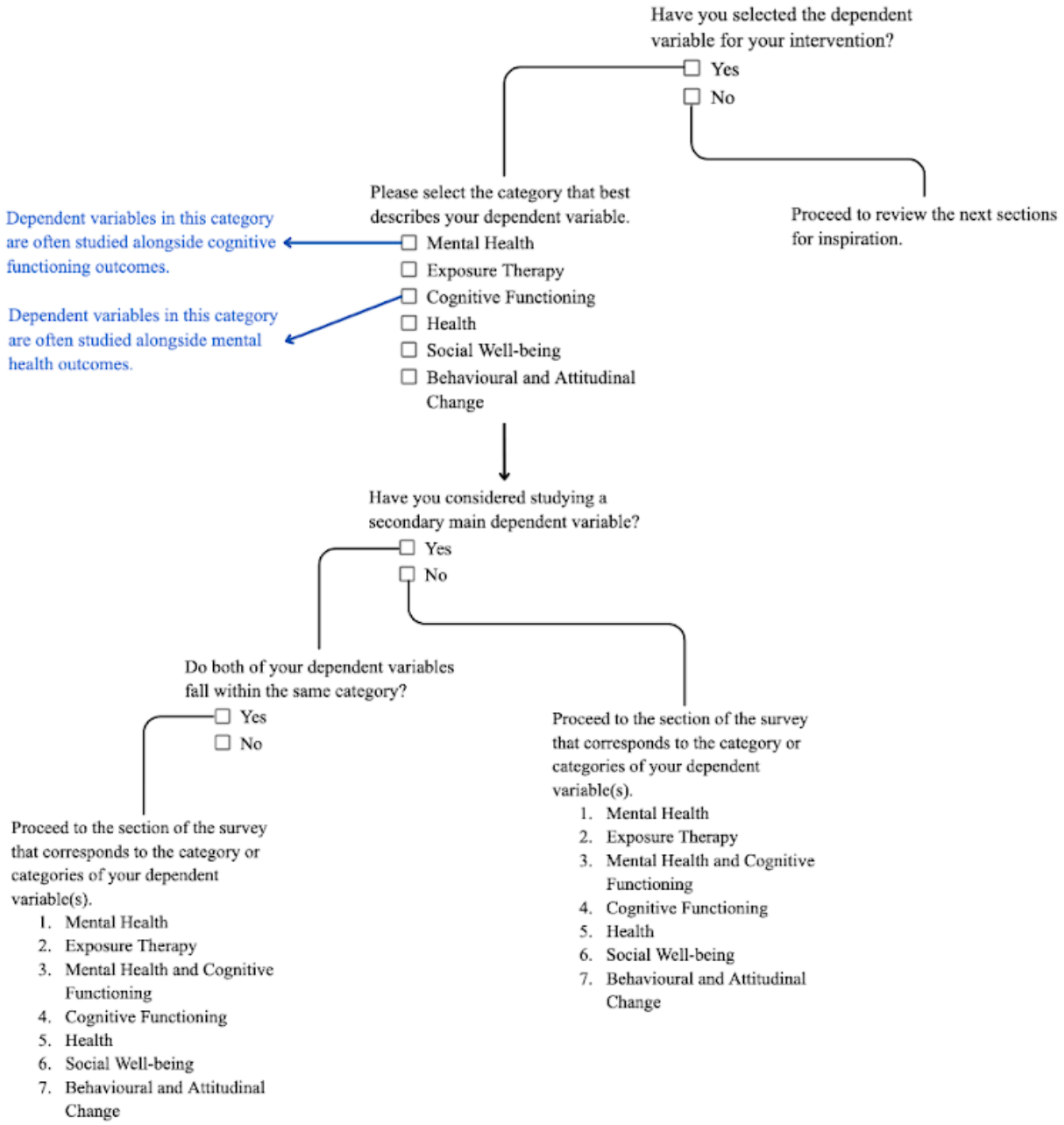
Psychophysiological Measures

1. Electrodermal activity (EDA): Measures the variation of electrical activity of the skin in response to sweat secretion, also indicative of autonomic nervous system activation (Benedek & Kaernbach, 2010).
2. Galvanic Skin Reaction/Response (GSR): Comparable to EDA.
3. Respiratory rate (RR)/ Respiration (RSP): Frequency of breathing, indicative of autonomic nervous system activity (Ritsert et al., 2022).
4. Self-assessment model (SAM): A non-verbal image assessment technology that directly measures the pleasure, arousal, and dominance associated with a person's emotional response to various stimuli.

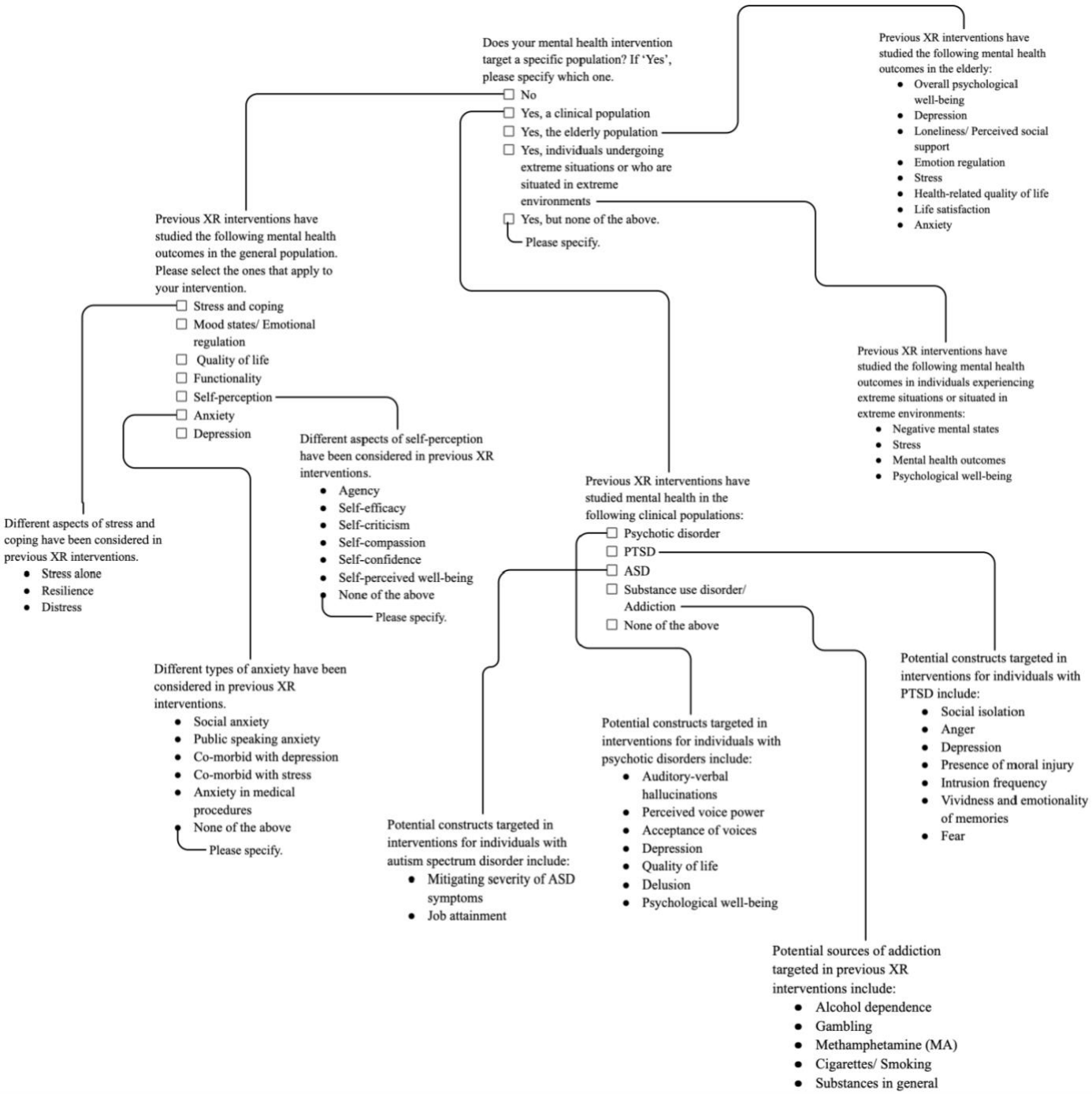
Appendix R

Checklist Mind Maps

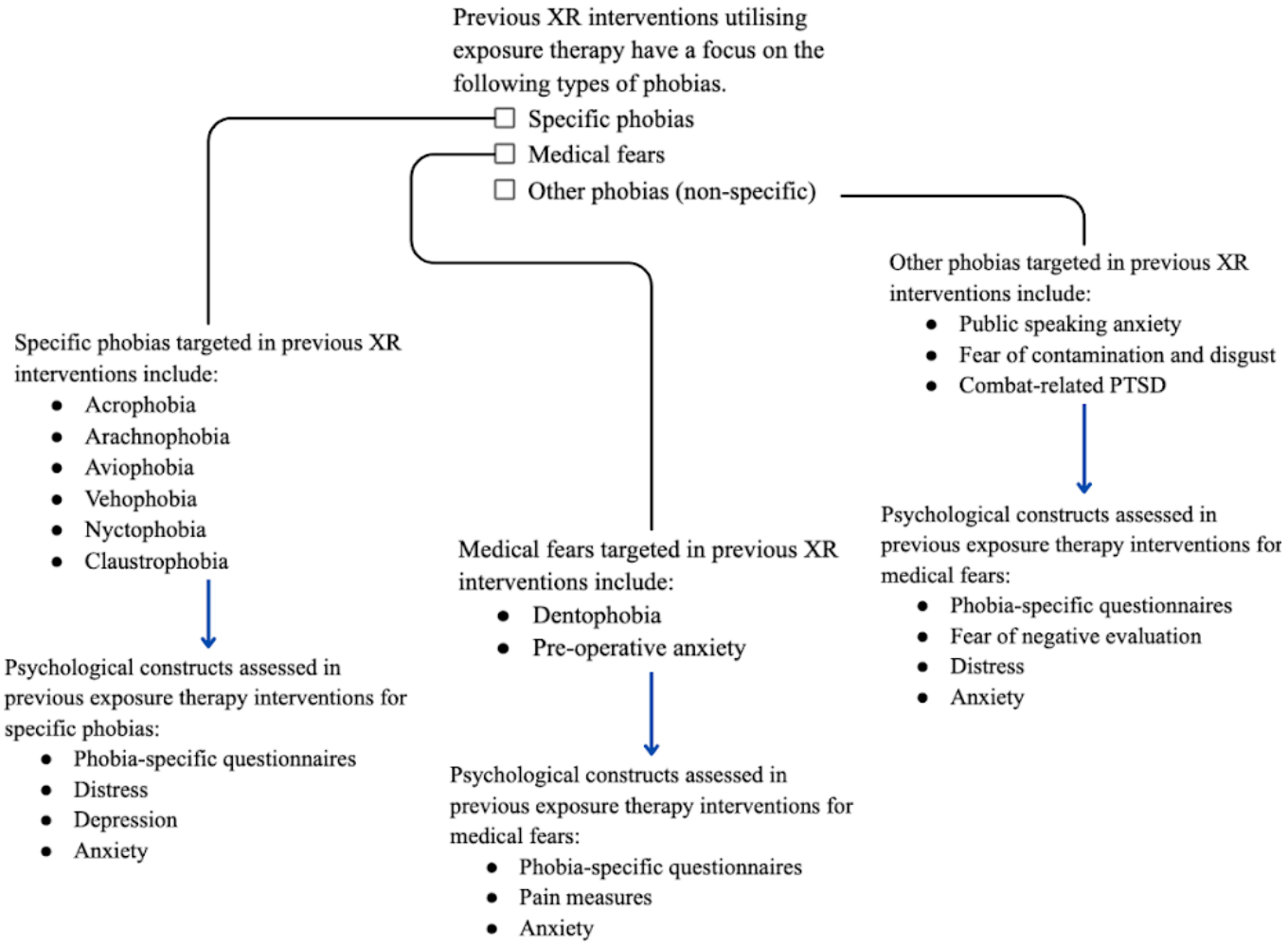
R1. Establishing the Main Category



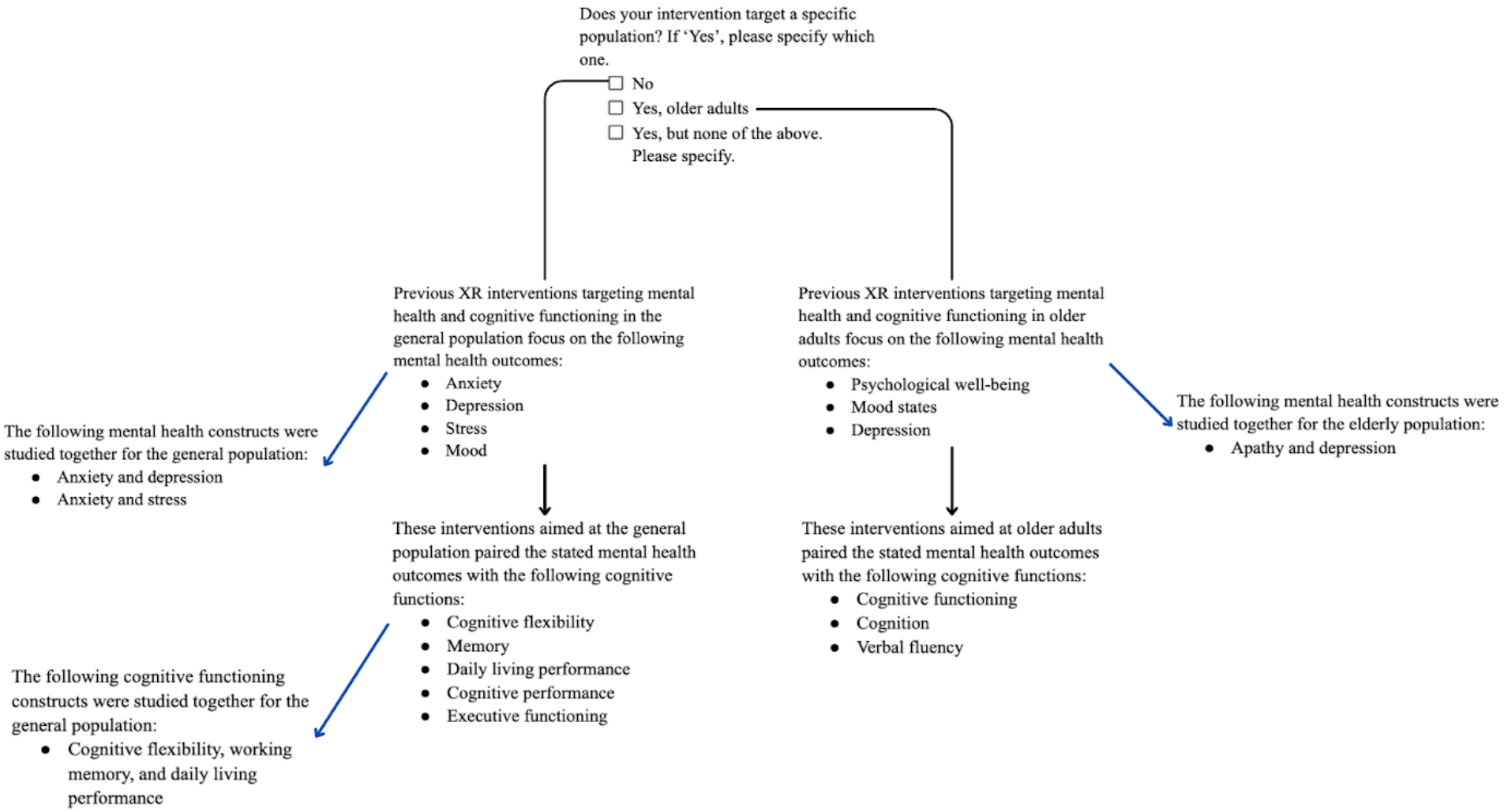
R2. Mind map for the “Mental Health” Section of the Checklist



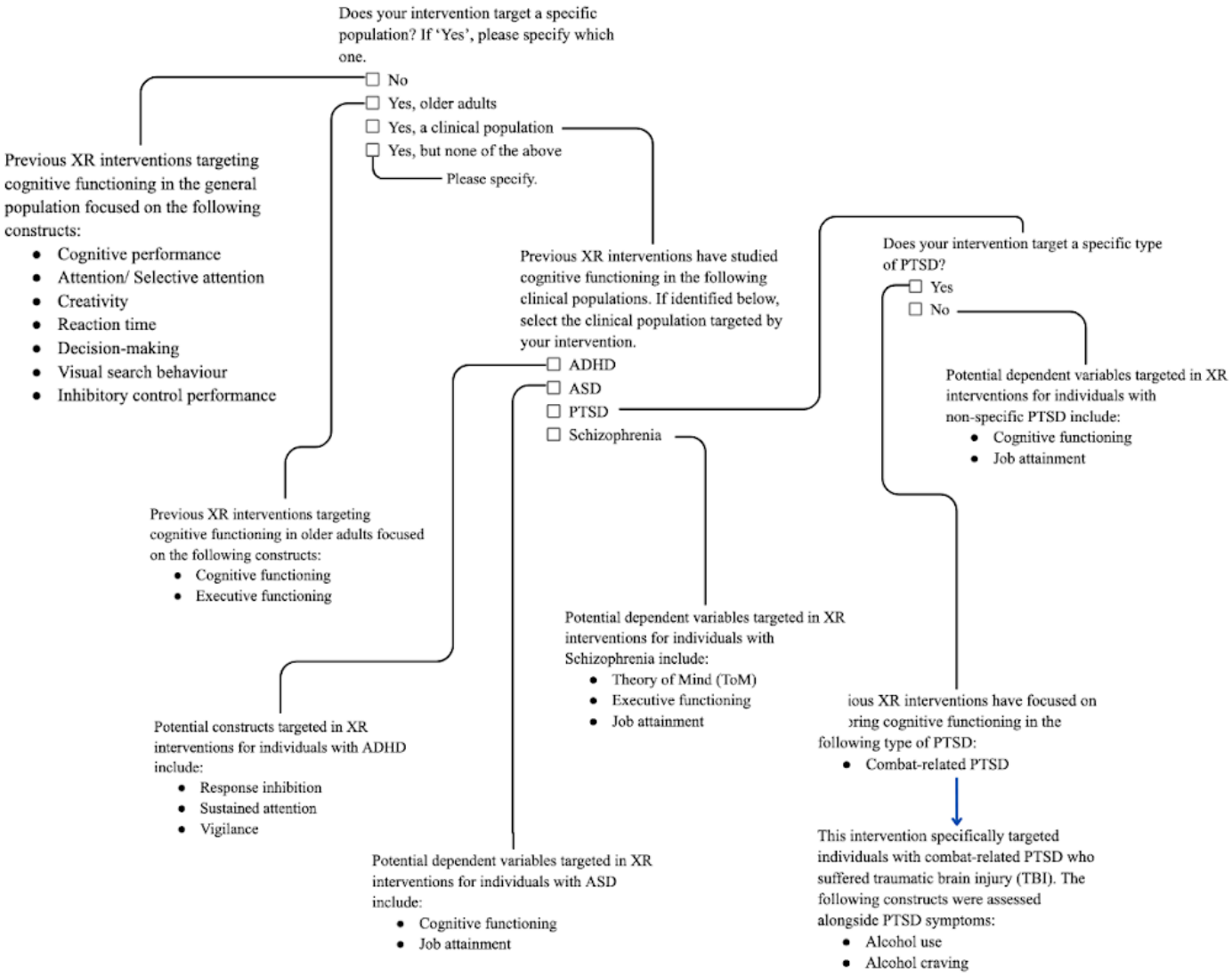
R3. Mind map for the “Exposure Therapy” Section of the Checklist



R4. Mind map for the “Mental Health and Cognitive Functioning” Section of the Checklist

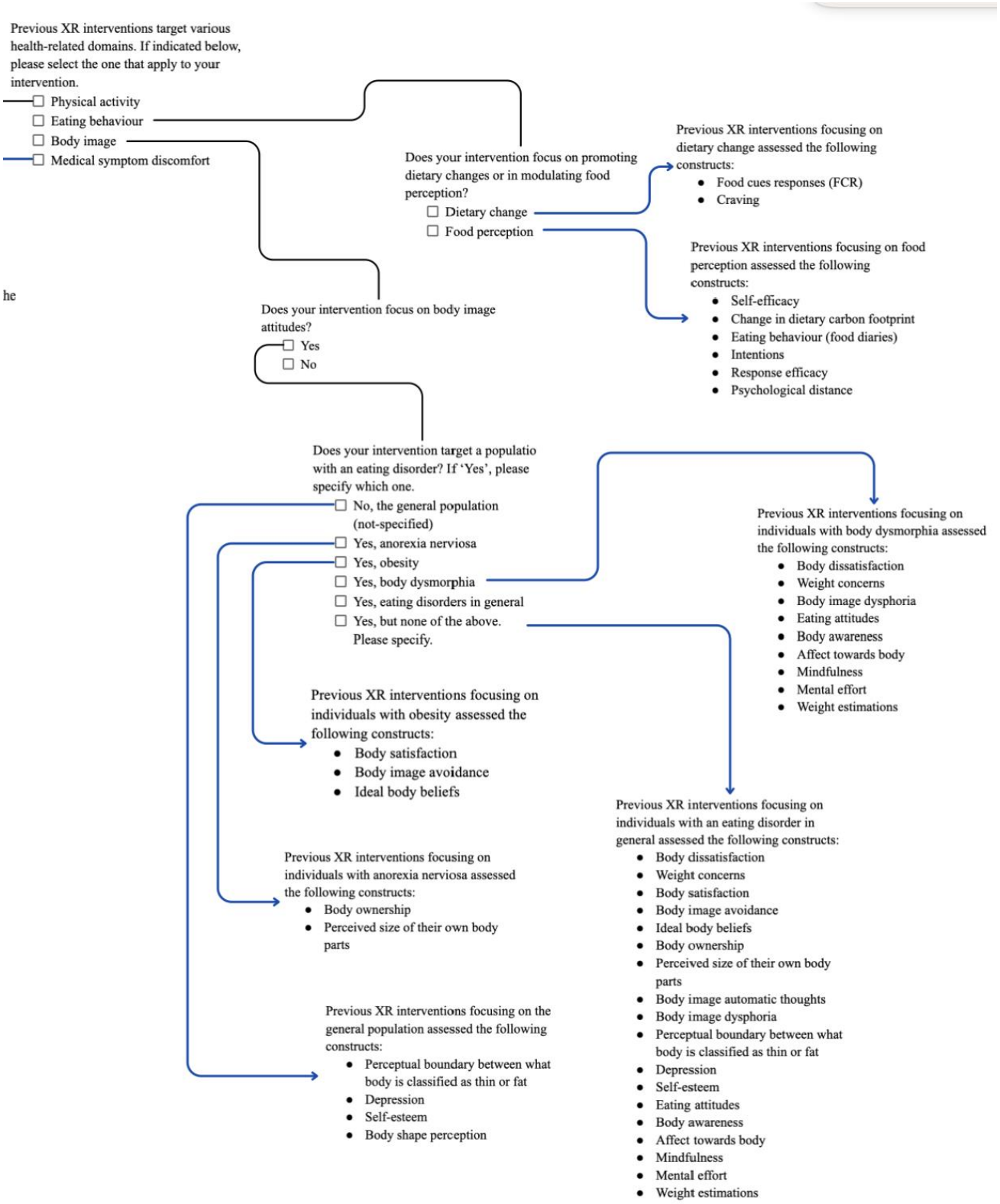


R5. Mind map for the “Cognitive Functioning” Section of the Checklist

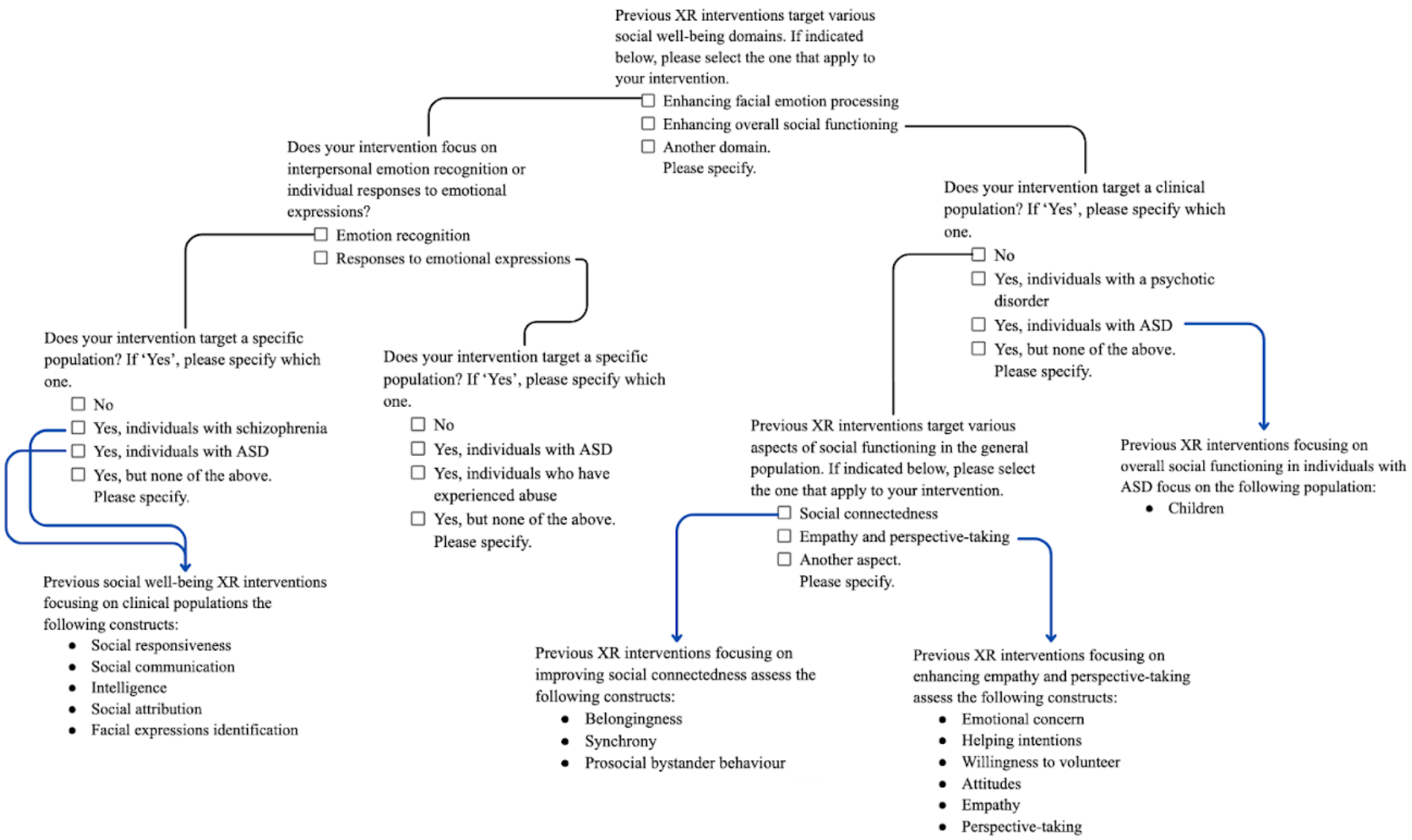
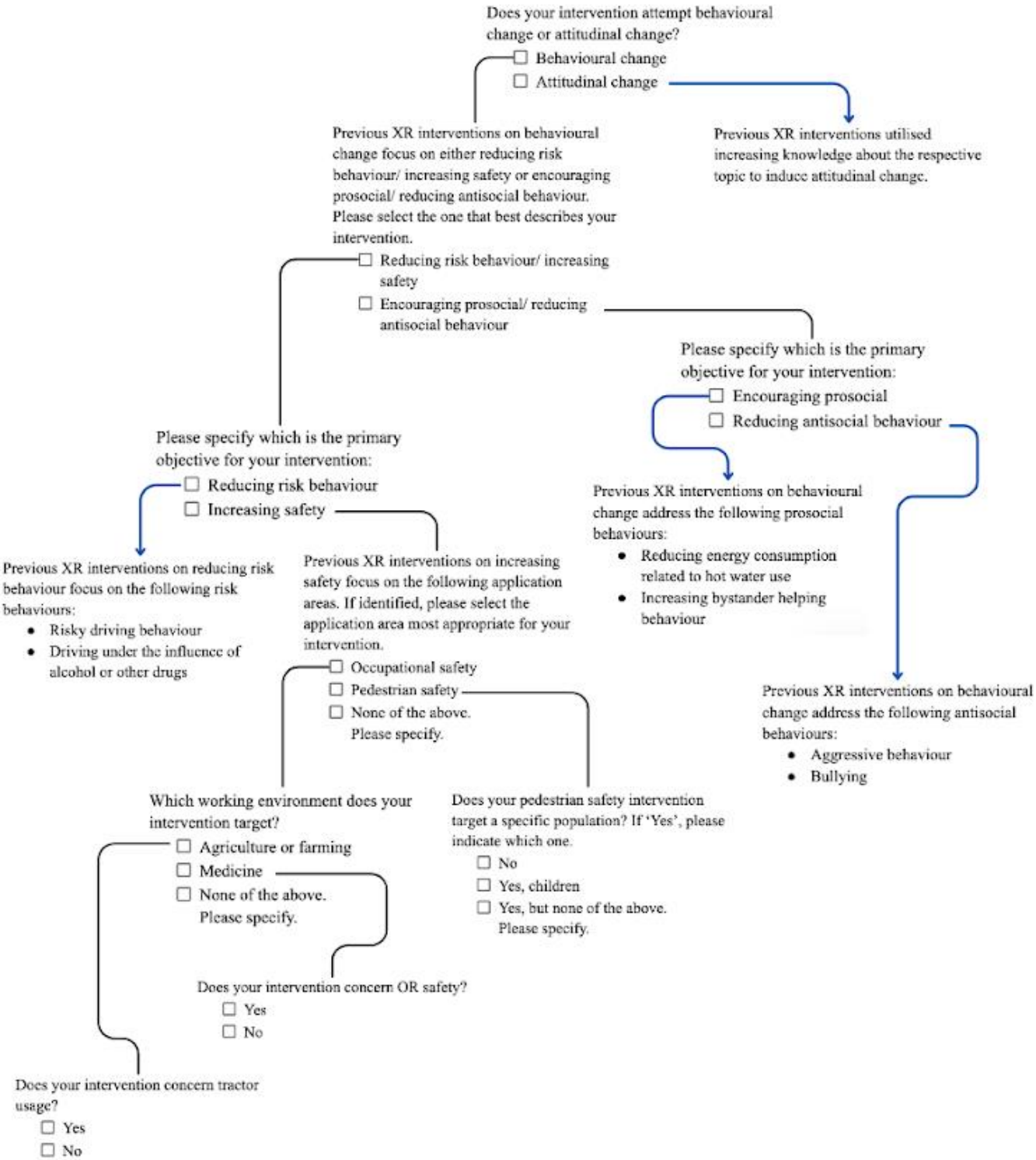


R6. Mind map for the “Health” Section of the Checklist (1/2)

R6. Mind map for the “Health” Section of the Checklist (2/2)



R7. Mind map for the “Social Well-being” Section of the Checklist



R7. Mind map for the "Behavioural and Attitudinal Change" Section of the Checklist