

STUDY PROTOCOL

Open Access



# A protocol for the development and validation of a virtual reality-based clinical test of social cognition

M Matre<sup>1,2\*</sup>, T Johansen<sup>1,3</sup>, A Olsen<sup>4,5,6</sup>, S Tornås<sup>1</sup>, AC Martinsen<sup>1,7</sup>, A Lund<sup>3</sup>, F Becker<sup>1,8</sup>, C Brunborg<sup>9</sup>, J Spikman<sup>10</sup>, J Ponsford<sup>11,12</sup>, D Neumann<sup>13</sup>, S McDonald<sup>14</sup> and M Løvstad<sup>1,2</sup>

## Abstract

**Background** Impairments in social cognition are common after traumatic brain injury (TBI) and may have severe negative consequences for patients and their families. Most tests of social cognition have limited ecological validity due to simplistic and contrived social stimuli with limited relevance to everyday social functioning. There is a need for measures of social cognition that reflect the dynamic, multimodal and contextualized nature of social situations and that predict real-world functioning. Three hundred sixty-degree (360°) Virtual Reality (VR) video can increase ecological validity through enhanced social presence, or a sense of “being there”. This paper describes the development and protocol design for validation of a Norwegian VR-version of The Awareness of Social Inference Test (TASIT), which is a widely used video-based test of social cognition.

**Methods** Development of VR TASIT included filming 61 short videos depicting social interactions in both VR and desktop format, using a 360° camera. Software for standardized test administration and collection of performance data was developed in *Unity*, for administration on both VR and desktop interface. The validation study will test the reliability and validity of VR TASIT in participants with TBI ( $n = 100$ ) and healthy controls ( $n = 100$ ). Half of the participants will perform the desktop version, and the other half the VR version. Analyses will include known groups validity, convergent and divergent validity, as well as test–retest reliability of VR TASIT. A comparison of the ability of TASIT VR and desktop versions to predict real-world functioning (ecological validity) will be explored using the *Social Skills Questionnaire for TBI* and *La Trobe Communication Questionnaire*. Finally, the levels of perceived social presence of the stimulus materials and prevalence of cybersickness after exposure to the virtual environment will be documented.

**Discussion** It is expected that VR TASIT will have comparable or better psychometric properties than the desktop version, and that the hypothesized increased level of social presence experienced in a virtual environment will result in improved ecological validity. More broadly, benefits and limitations of using VR video as stimulus material in assessment of social cognition and considerations for future development and clinical validation are discussed.

\*Correspondence:

M Matre

martin.matre@sunnaas.no

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Trial registration** The study protocol was pre-registered in ClinicalTrials (April 4<sup>th</sup> 2022, NCT05309005). The study was retrospectively registered in Open Science Framework (December 15<sup>th</sup> 2022, osf.io/2vem8).

**Keywords** Virtual reality, 360-degree video, Social cognition, Assessment, Traumatic brain injury, Rehabilitation

## Background

Social cognition refers to the ability to identify and interpret social cues in order to make sense of social situations and respond appropriately [1]. Social cognitive impairments are common after traumatic brain injury (TBI). For instance, 13–39% of people with moderate to severe TBI show impaired ability to recognize emotion in facial expressions [2]. Social cognitive impairment is a leading cause of social isolation, relationship disintegration and unemployment in this population [3, 4].

Despite the severe negative consequences for patients and their families, social cognition is rarely assessed systematically by clinicians: In a survey of 443 clinicians, 84% reported that more than half of their patients with severe TBI had social cognitive impairments, but 78% acknowledged that they infrequently or never assessed this domain with standardized assessment tools [5]. The most frequently reported reason for this was the lack of access to standardized assessment tools with relevance to everyday social functioning, i.e. sufficient ecological validity to be clinically relevant. The lack of ecologically valid tests of social cognition has also been addressed by leading researchers in the field [6–9].

Social cognition is an umbrella term involving several related domains, including the ability to recognize emotions in others, inferring other people's state of mind, taking the social context into account and regulating social behavior [10]. The social cognitive domains that have received the most attention in TBI research are empathy, emotion recognition and Theory of Mind (ToM) – the ability to take another person's perspective [11, 12]. A recent scoping review [12] found that the most commonly used stimulus materials in research on impaired emotion recognition after TBI are the Ekman and Friesen photographs [13, 14]. Here, emotional expressions are conveyed by actors from the 1970s in black and white. Measures of ToM typically present participants with very short stories in the form of cartoons or short text vignettes, asking them to interpret what is implicitly communicated by one of the actors in the story [15]. These stimuli are designed to minimize the effect of potential confounding variables, thus increasing internal validity in controlled experiments. However, the primary clinical concern is to predict patients' behavior in everyday social situations, i.e., ecological validity. Several studies have found that tests of social cognition developed for research purposes have limited value [16, 17], despite the

impairments reported by both clinicians, patients with TBI and their relatives [18].

Everyday social cognition depends on many sources of information, e.g. facial expression, verbal content, body language, tone of voice and context [9, 19]. Furthermore, social information unfolds and changes over time and is embedded in a specific context [20]. Tasks that incorporate naturalistic stimuli that are dynamic, multimodal, and context-embedded, may increase generalizability of performance on social cognitive tasks to everyday social situations. This would mean moving beyond stimuli such as photographs and text vignettes, as well as adding background information usually available when interpreting social situations. However, few such tests are available for clinicians today.

One example of a test with dynamic and multimodal stimuli is The Awareness of Social Inferences Test (TASIT), which uses videos of everyday social situations to measure emotion recognition and Theory of Mind [21]. It assesses emotion perception and Theory of Mind, as the test person is asked to interpret the beliefs, intentions, and emotions of people in everyday social situations. TASIT performance predicts everyday social functioning in TBI [22], likely as a result of the increased social presence afforded by the stimulus materials. However, watching videos on a two-dimensional screen affords limited social presence [23], i.e. the sense of actually being present in the social situation. In real life, social cognitive impairments manifest themselves in social situations that patients are part of. The lack of this dimension reduces the ecological validity of TASIT.

Virtual Reality (VR) technology is well suited to generate realistic stimuli that can generalize to everyday social situations. VR can be defined as an “externally mediated presentation of sensory stimuli that enables the person to perceive an artificial environment as non-synthetic to a greater or lesser extent” [24]. A head mounted visual display obscures the external environment, which together with audio input allows for immersion in the virtual environment. VR software using stimuli similar to real world cues is effective in both assessment and treatment of many mental disorders, including anxiety disorders, eating disorders and alcohol and substance use disorders [25]. A likely reason why VR is successful in both predicting and treating real world phenomena is that it facilitates a sense of presence, i.e. “the perceptual illusion of non-mediation that occurs when a person fails to perceive

the existence of a medium” [26], which is not attainable in a two-dimensional medium. Social presence refers to the sense of being with another person [27]. A range of factors influences social presence, from the presence or absence of a visual representation of the other person to the photographic and behavioral realism of that person [28]. VR has other benefits in addition to increasing ecological validity through enhanced social presence. As VR technology allows for standardized stimulus presentation, the internal validity of the test can be preserved. Furthermore, VR offers time- and cost-efficient automated test administration and recording of responses [29]. A minority of VR users experience adverse effects, such as headaches or nausea, referred to as cybersickness [30]. Both clinical practice and preliminary research indicates that persons with TBI in the chronic phase tolerate the use of VR well, but there are few empirical studies on cybersickness in the TBI population [31].

### Aims and objectives

Our long-term goal is to establish an ecologically validated measure of social cognition for patients with TBI. To this end, the primary aim of our overall study is to develop a Norwegian VR-version of the TASIT (VR TASIT) and explore its psychometric properties, including ecological validity, in participants with and without TBI. For the present paper, our objective is to describe the development and design for future validation of the Norwegian VR TASIT through:

- 1) A detailed account of the development of VR TASIT.
- 2) Description of the protocol for validation of VR TASIT, i.e., procedures for determining.
  - A) construct validity and test–retest reliability of VR TASIT
  - B) ecological validity of VR TASIT
  - C) perceived level of presence in VR TASIT.
  - D) occurrence of cybersickness, i.e. symptoms of physiological discomfort, after performing VR TASIT.

## Methods

### Development of VR TASIT

The original TASIT [21] consists of three subtests, one measuring emotion recognition, the Emotional Evaluation Test (EET), and two measuring Theory of Mind, the Situational Inference Test—minimal (SIT-m) and the Situational Inference Test—enriched (SIT-e). In SIT-m, the characters are either sarcastic or sincere; in SIT-e they are either sarcastic or lying. After each EET video, participants are asked to choose among seven emotion

categories (happy, angry, sad, anxious, surprised, revolted or neutral). After each SIT-m/SIT-e video, four yes/no-questions are presented, assessing the ability to infer a character’s underlying belief, intention, emotion and meaning. Each subtest consists of videos of 15–60 s duration and contains 1–3 actors in various social scenes. It has an A and a B form for retest purposes, resulting in a total of 118 videos. Each form takes approximately 60 min to administer. Responses to the multiple-choice questions are recorded on a pen and paper form after each video. The stimuli are designed to be natural and unambiguous, causing ceiling effects for healthy participants and reduced and more variable scores for persons with TBI [32]. It has been demonstrated that TASIT has good construct validity and test–retest reliability [33]. TASIT performance has been shown to be affected in a range of clinical populations with impaired social cognition, including schizophrenia, [34], frontotemporal dementia [35], and TBI [36, 37].

### Planning and preparations for production of VR TASIT

It was decided that VR TASIT should track the original test as closely as possible except for level of social presence, i.e. the VR aspect. The overall format of the original test was preserved, as was item order instructions, dialogues, as well as questions and answers. For practical purposes, it was decided to produce the three subtests that make up the A Form and not the alternate B Form, as the former is most frequently used in the research literature [32]. A collaboration was established with Prof. Skye McDonald, the researcher who developed the original version of TASIT. McDonald has taken part in several meetings and discussions throughout the production process.

### Development of stimulus material

The A Form of TASIT consists of 61 videos in total, 59 test items and two practice items. Prior to the video production, clinical neuropsychologists (authors MM and ML) examined the original videos in order to determine the need to adjust the content to preserve face validity, as the original videos were filmed in Australia in the early 00 s. Some cultural differences were anticipated, but as none emerged after examination of the videos, it was decided not to make any changes based on culture. However, as two decades had passed since TASIT was produced, some modernization was needed. For example, as most purchases in Norway are presently made digitally or with credit cards, cash is seldom seen and scenes with coins or notes were adapted somewhat. In addition, landline telephones were replaced with mobile phones. The actors’ appearance and the locations were a natural reflection of present-time Norway and differed from

the original test. Only one video was replaced, for modernization purposes. In SIT-e, task 10 a man is teased for being overweight. This was replaced with a scene with identical content, but overweight was replaced with looking tired after having spent the night out partying, as this was considered more culturally acceptable.

It was decided that the actors should from time to time “break the fourth wall”, the invisible wall between the actor and the audience/viewer [38], by looking directly at the camera (see Fig. 1). This is a deviation from the original TASIT, where actors never gaze into the camera. This decision was made to take maximum advantage of the higher level of social presence in the virtual reality medium, enhancing the participants’ sense of being part of the social situation [39].

All dialogues were translated from English to Norwegian without major changes. English names were replaced with Norwegian names. While verbatim translation was strived for, slight alterations had to be made in some videos to preserve the intended meaning of the original. For instance, sentences that began with the word “well” in the original dialogues were replaced with a Norwegian word with a different literal meaning, while serving the same pragmatic function.

The original TASIT videos alternate between using a neutral black background and studio sets (office, kitchen,

etc.). In order to maximize virtual reality’s propensity for presence [40, 41], and thus increase the ecological validity of the stimuli, it was decided to film all videos in settings where social interaction naturally occurs, such as private apartments and in various public places (Fig. 2).

### Filming

A professional film producer was hired for filming and editing of the 61 videos. In addition, the producer was responsible for hiring actors, securing appropriate locations and logistics related to filming. The importance of an even distribution of the actors’ gender, age and ethnicity, as well as realistic social contexts, were conveyed to the producer. As the expression of emotions and beliefs in the TASIT videos were designed to be simple and clear for neurotypical people with average social skills [21], several steps were taken to ensure that the actors understood the importance of expressing emotions and beliefs in an exaggerated, yet natural, style and to as far as possible express one emotion only in each scene. Before filming, the rationale of TASIT was explained in detail to the film producer by specialists in neurorehabilitation (authors M.M. and M.L.) and on the first day on set, a clinical neuropsychologist (author M.M.) was present to brief the actors about the purpose of the production. The producer instructed the actors to convey social cues unambiguously throughout filming, and first author M.M. and the producer collaborated very closely throughout the entire production, which is considered an important asset of the process.

Filming commenced in August 2021 and finalized in March 2022. Five days of filming were required to film the 61 videos. In all, nine different locations were used. Two were private residences, the rest public settings (cafe, public library, office building, and hospital). At all sets, different rooms and spaces were used to maximize novelty. A clinical neuropsychologist (M.M.) was present on set to ensure that the actors performed in accordance with the requirements of TASIT. Before each scene was filmed, the actors were told the question(s) participants



**Fig. 1** Example of an actor turning to the camera to enhance the level of social presence



**Fig. 2** To the left: A screenshot from an emotion recognition video in the original TASIT. To the right: A screenshot from the corresponding video in VR TASIT

would be asked after watching the scene, as well as the correct answers.

A GoPro MAX 360 Action Camera, a 360-degree camera that captures the full circle of the horizontal plane of the surroundings, was used to film all scenes (Fig. 3). Compared to a lens that is limited to capturing e.g. 40–60°s of any given field of view, a 360° video immerses the viewer in a realistic virtual environment [42]. Displaying the 360° videos on a head mounted display, which occludes external stimuli and provides additional depth, increases levels of presence beyond what is experienced when watching a desktop version of the same videos [43]. The raw material of each take was reviewed on set by both producer and author MM immediately after filming each item, to ensure that the intended content was achieved. If not, a new take of the scene was performed.

### Editing process

In the intervals between filming, raw materials were edited, using Final Cut Pro X software. Most videos consisted of one scene only, except for the eight items in SIT-e that have a prologue or epilogue providing participants with a cue to help participants infer the actor's true belief. Thus, relatively little editing was required. While all videos were recorded with the 360-degree camera, two versions of each video were produced, one VR version in a format compatible with commercially available VR equipment and one in standard 2D desktop format.

### Postproduction expert considerations

The videos were reviewed by an expert panel consisting of three persons with extensive clinical and research experience within brain injury rehabilitation (authors M.L., S.T., T.J.). The purpose was to ensure that the content of the videos conformed to the original TASIT in terms of the emotions, beliefs and intentions expressed by the actors, as well as a general quality assessment, i.e. not validity testing of the entire test as such. The review consisted of a group administration of TASIT, where the panelists gave their response after having viewed each video, without knowledge of the other panelists' responses or knowledge of the correct responses.

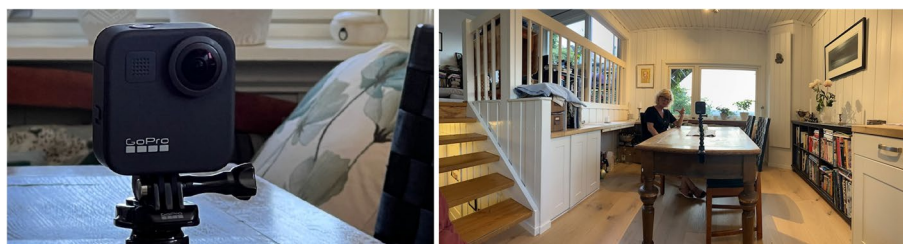
For 87,5% of all questions (a total of 155 questions across the 61 videos), at least two of the three experts' responses were correct. No items were given an incorrect response by the entire expert panel. For EET, there were no items where two of the three answered incorrectly. In SIT-m, there were two questions (out of 60) that two of the three experts answered incorrectly. In SIT-e, there were 14 (out of 64) questions where two of the three experts answered incorrectly. On a positive note, these results are largely in line with the scores of healthy controls in the original TASIT [21]. Still, some quality issues with a subset of videos were addressed.

After the panelists had provided their responses, the videos were scrutinized qualitatively, both in terms of the acting performance and if there were issues with the location. One example of the latter was in a video where an actor pointed to and talked about a car, and it was identified that rotating the head 60° in VR would reveal that there were no cars there to be seen. In total, issues were identified in eight videos. In five of these, a majority of the panel found that either sarcasm or an emotion was poorly expressed by the actor. In three videos, problems were identified with the location and/or the 360° presentation. A further 15 videos were identified as potentially problematic, either because two of the three panelists had given an incorrect response to one or more of the four questions in SIT-m and SIT-e or because of minor issues with the actors performance.

It was decided that the eight videos identified by the panel as problematic should be shot again. In addition, the 15 videos that were identified as potentially problematic were shown to a panel of non-experts, consisting of 10 healthy individuals. No limit was set as to how many errors were acceptable, instead a combination of qualitative and quantitative reasons guided which videos to shoot again, resulting in an additional six videos to be reshot.

### Development of digital test instructions

The test instructions were translated into Norwegian by author MM in collaboration with author ML. To familiarize participants with the virtual environment it was



**Fig. 3** To the left, the 360-degree camera used to film all videos. To the right, film set example, with all on set except for the actor out of the camera's field of view. A written permission to display the camera and company logo was obtained from the copyright holder

decided to record a VR video with a virtual test administrator delivering the first introductory test instructions, with a duration of approximately 1 min. The remaining instructions were delivered with text and audio, but not filmed.

**Software development**

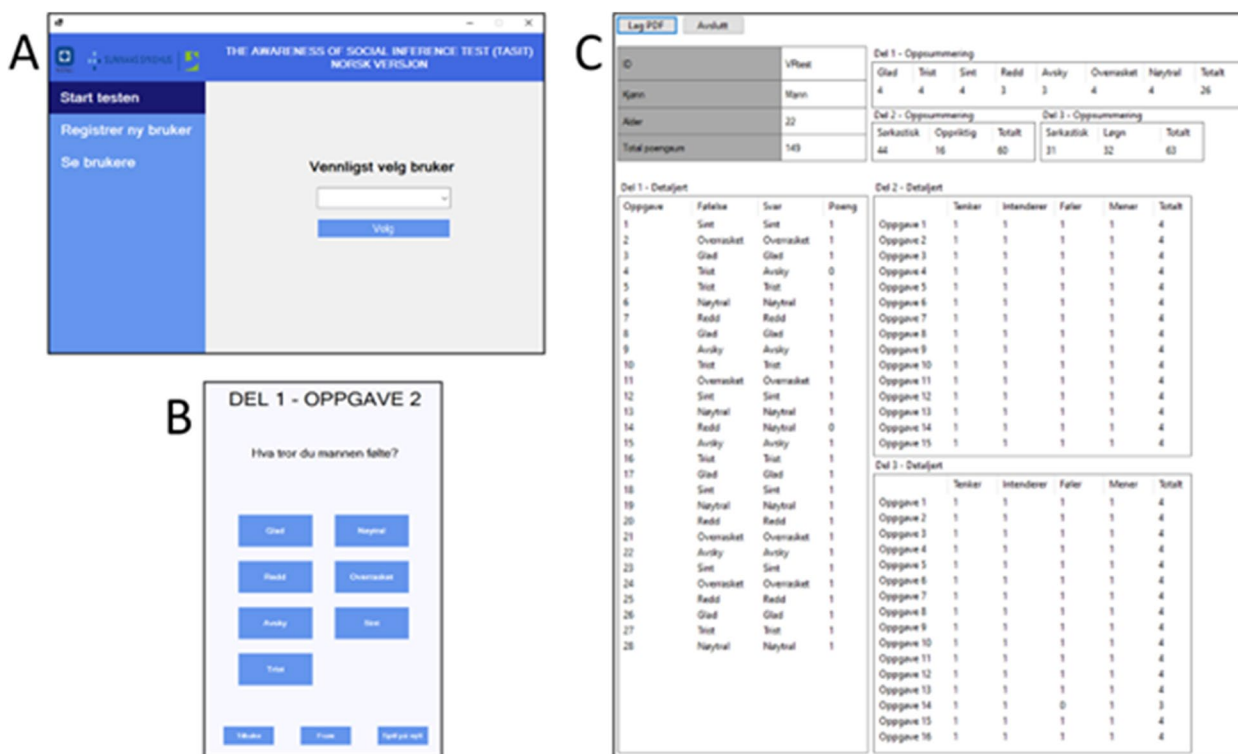
A computer program was developed in order to create unique users (i.e., participants), administer the test and generate a score sheet for each user. The rationale for this was to save administration resources, as the original TASIT requires a test administrator to query the participant after each video for a response and then record the answers on paper, as well as to avoid errors in registration of responses and calculation of total scores.

Conceptual discussions were had with the software developer to convey the necessary software functions and structure. It was decided to create a menu-based VR computer program (Fig. 4) with the following functions:

1. *Create new user.* Unique users are registered in a menu. For the initial version, only “name/ID”, “age”

and “gender” are recorded, this could be expanded on in a future version of the test.

2. *Administration of TASIT.* Options are to administer the whole test or to select one or two of the subtests. If the test is aborted midway throughout administration, it picks up at the same place when run again. Each item begins with a prompt (e.g., “Focus on the man to the left”) and a box to be selected for the video to start. When the box is selected, a 3–2–1 countdown appears, to signal that the video is about to begin, followed by the video. After the video, the item’s question(s) appears on the screen. For SIT-m and SIT-e, each of the four questions appear sequentially. Below each question, boxes represent response alternatives to be selected by the participant. When a response is selected, the program moves on to the next question in SIT-m and SIT-e or to the next item in EET/after the fourth question in a SIT-m/SIT-e task.
3. *Results.* As new users are created, they are added to a list of all users in a separate submenu. When a specific user is selected, a result form is accessed, where both item level results and total scores for each sub-



**Fig. 4** **A:** Software main menu, with options to either select an existing user to start a new test session, register a new user or to show a list of all registered users and create a score sheet. **B:** Example of response screen, which appears after each video has finished. The header shows which part and task the participant is currently at, the question indicates which actor to respond to and the big blue boxes represent the alternative emotion categories. The bottom options are to see the video again, move forward or move backward in the test. **C:** Example of score sheet. Background information in the top left corner, summary scores in the top right corner and item level scores underneath

test are displayed. The result form can be exported to pdf-format or printed.

The software was implemented in C# with object-oriented programming principles, divided into separate classes for different parts of the program. For the desktop version, a Windows Forms application was developed in Visual Studio 2019. The VR version was an extension of the desktop version, developed in Unity. The program was implemented with a modified model-view-controller architecture, with view and control combined as one component. Windows Forms framework was used to create the user interface, which handles user operation, updates models, and displays relevant data in the application. Throughout the development process, principles from user-centered design were implemented, frequently testing the applications on users to ensure that it was user-friendly.

Throughout the stages of software development, it was tested and reviewed by author MM to ensure both the usability of the functions and that the software conformed to the structure of the original TASIT. The software was then tested on both rehabilitation professionals familiar with TASIT and patients with TBI, to confirm that the program was user friendly for both

administrators and participants, that it performed as intended and to eliminate minor errors.

### Validation study protocol

#### Study design

The study is a prospective observational cohort study. Patients will be randomly assigned to perform TASIT in either VR (VR TASIT) or 2D version (2D TASIT). All participants, regardless of TASIT-condition, will report on measures determining validity both at baseline (before randomization) and at T2 16 weeks later. An equal number of healthy adults will be matched to the patient group with respect to age, gender, and education and perform either VR or 2D TASIT (see Fig. 5).

#### Settings and study population

Data collection will take place at the Sunnaas Rehabilitation Hospital (SRH) VR-laboratory from November 2022 to spring 2024. SRH is a tertiary level rehabilitation hospital that treats approximately 1000 patients with acquired brain injury each year. We will recruit former inpatients at SRH with moderate to severe TBI.

*Participants needed to fulfill all of the inclusion criteria noted below:*

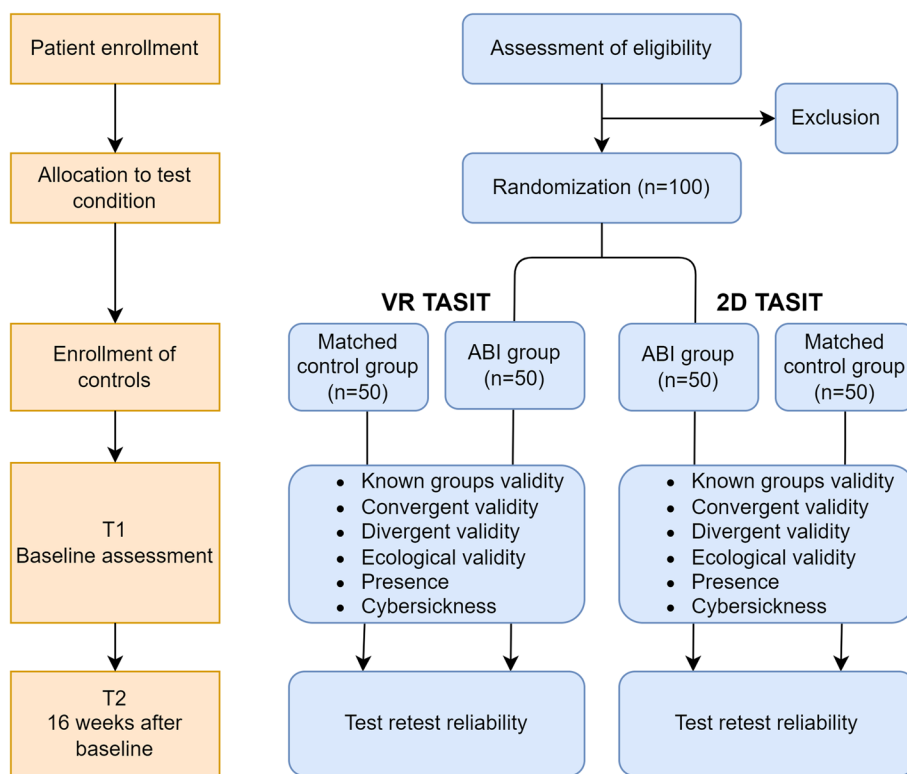


Fig. 5 Study design

- A) Moderate or severe TBI according to the diagnostic criteria of The American Congress of Rehabilitation Medicine: Duration of loss of consciousness > 30 min, a Glasgow Coma Scale score < 13, or post-traumatic amnesia > 24 h [44].
- B) Positive trauma-related intracranial findings on CT and/or MRI.
- C) Minimum 12 months after injury and maximum 10 years after injury.
- D) 18–65 years of age.
- E) Physically able to operate VR-equipment.
- F) Able to understand instructions in Norwegian.
- G) Cognitively capable of providing informed consent.

*Exclusion criteria were:*

- A) Language impairments affecting the ability to understand instructions.
- B) Motor impairments affecting the ability to utilize the VR equipment.
- C) Visual neglect.
- D) Severe mental illness.
- E) Comorbid neurological disorders.
- F) Non-Western cultural background.

Identical inclusion and exclusion criteria apply to the healthy control group, except for a prior history of TBI.

Self-report and informant questionnaires will be collected digitally, by means of a secure platform for data collection and storage (Service for Sensitive Data, TSD). TSD is an IT-platform at the University of Oslo with a secure server approved for storage of sensitive research data. Data collections were handled by questionnaires created with nettskjema.no, a survey solution developed and hosted by the University of Oslo [45].

**Validation measurement**

The construct validity, test–retest reliability, and ecological validity of both TASIT versions will be investigated. In addition, any adverse effects of exposure to VR TASIT (i.e., cybersickness) and the participants' experienced level of social presence is assessed (See Table 1 for an overview of measurements).

**Construct validity**

The construct validity of both TASIT versions will be established if it is demonstrated that they (1) discriminate between two groups known to differ on a measured construct (known groups validity), (2) correlate with other tests of social cognition (convergent validity) and (3) do not correlate with tests that measure general cognition (divergent validity).

**Table 1** Measurements used for validation

MEASURE	TIME	
	T1	T2
<b>Questionnaires</b>		
Interpersonal Reactivity Index, IRI [50]	x	x
Social Skills Questionnaire for Traumatic Brain Injury, SSQ-TBI [62]	x	x
LaTrobe Communication Questionnaire, LCQ [63]	x	
Behavioral Rating Inventory of Executive Functions—adult version, BRIEF-A [54]	x	
Simulator Sickness Questionnaire, SSQ [67]	x	x
Multimodal Presence Scale, MPS [66]	x	x
General Anxiety Disorder 7, GAD-7 [57]	x	x
Patient Health Questionnaire, PHQ-9 [58]	x	x
<b>Social Cognitive tests</b>		
Emotion Recognition Task, ERT [48]	x	x
Hinting Task [46]	x	x
<b>Cognitive tests</b>		
CPT III – Hit Reaction Time and Standard Deviation of Hit Reaction Time [52]	x	
WAIS IV: Coding [51]	x	
WAIS IV: Similarities [51]	x	
WAIS IV: Matrices [57]	x	
WAIS IV: Digit Span Backwards [51]	x	
WAIS IV: Digit Span Sequencing [51]	x	
D-KEFS: Color Word Interference Test 3 [53]	x	
D-KEFS: Trail Making Test 4 [53]	x	



### **Known groups validity**

Comparison of the performance of participants with TBI and healthy controls in both the 2D and VR TASIT versions (total score and score on each subtest) will be performed, to assess whether the VR version is superior to the 2D version in discriminating between social cognitive impairment and normal performance.

### **Convergent validity**

Performance on TASIT will be compared with performance on established tests of three social cognitive domains: Theory of Mind, emotion recognition and empathy. The Hinting Task is a measure of Theory of Mind that assesses understanding of people's intentions from indirect messages [46]. The task consists of 10 text vignettes of a protagonist expressing an indirect message to another person. Participants are asked to describe the meaning behind the indirect messages. The Hinting Task has been translated to Norwegian and validated in Norwegian patients with schizophrenia [47], but not in patients with TBI. The Emotion Recognition Task (ERT) measures emotion recognition by asking participants to label facial expressions from photographs [48]. The ERT has well-established psychometric properties and correlates with performance on the original TASIT [49]. The Interpersonal Reactivity Index is a self-report measure of empathy [50]. It contains 28 items that are answered on a 5-point Likert scale.

### **Divergent validity**

Coding from WAIS IV [51] and Hit Reaction Time on the Conners' Continuous Performance Test 3rd edition (CPT III) [52] will measure processing speed. Sustained attention will be measured with the coefficient of variation (Standard deviation of Hit Reaction time / Hit Reaction time), where the final three test blocks will be compared to the first three. The mean scores on Backwards Digit Span and Digit Sequencing tests from WAIS IV will be used to measure working memory [51]. Executive functions will be assessed with Trail Making Test 4, a test of mental flexibility and Color Word Interference Test 3, which measures inhibition, both from the Delis-Kaplan Executive Function System test battery [53]. Everyday executive functioning will be assessed with the patient and informant versions of the Behavior Rated Inventory of Executive Functioning – Adult (BRIEF-A) [54], and abstract reasoning with Similarities and Matrices from WAIS IV [51]. As VR TASIT is both complex and dynamic, it is expected to correlate weakly with other cognitive functions, but not to overlap to a large degree. As mood disorders may

impair social cognitive functioning [55, 56], self-report measures of anxiety (Generalized Anxiety Disorder 7 (GAD-7) [57] and depression (Patient Health Questionnaire (PHQ-9) [58] are also included.

### **Reliability**

At T2, i.e., 16 weeks after T1, the two patient groups will perform the same TASIT version a second time, to determine the test–retest reliability of the two tests. The expected near ceiling effects of controls limits the ability to calculate reliability estimates in this population. The stability of social cognitive impairments over time [59], together with the inclusion criterion of minimum 12 months post TBI, justifies a relatively long test–retest interval and reduces the risk that recollection from T1 assessment interferes with performance on T2.

### **Ecological validity**

As research on social cognition after TBI is a relatively new field, no gold standard test exists against which the ecological validity, i.e., the relevance to social functioning in everyday life, of VR TASIT can be tested. Some measures of social skills that have been developed for other populations have been used in TBI samples, such as the Katz Adjustment Scale [60, 61], but these include psychiatric symptoms that are not relevant after TBI. The Social Skills Questionnaire after Traumatic Brain Injury (SSQ-TBI) is however promising, as it assesses informant-rated behaviors that are important for normal social interactions, as well as those impaired following TBI, such as emotion recognition, empathy, egocentrism and communication [62]. The SSQ-TBI taps 16 desirable and 24 undesirable behaviors, which yield negative and positive subscales, respectively. A final item measures a global evaluation of social functioning. The SSQ-TBI has been translated to Norwegian and a new informant version with identical items to the self-report version is incorporated into the protocol. The SSQ-TBI is relatively new, and empirical investigations are few. We will therefore also include the La Trobe Communication Questionnaire (LCQ) as a measure of ecological validity [63]. LCQ measures impairments in social communication with 30 items being rated by patients and informants. The LCQ has been translated into Norwegian [64] and discriminates between people with brain injury and healthy adults [65]. Ecological validity will thus be determined by how well both 2D and VR TASIT results correlate with a measure of everyday social skills (SSQ-TBI), and social communication (LCQ), as rated by patients and their close relatives.

### Assessment of social presence

The Multimodal Presence Scale measures the perceived physical, social and self-presence in a mediated experience on 15 five-point Likert-type questions [66]. It has been translated to Norwegian and will be used to establish if differences between scores in the two TASIT versions are associated with differences in perceived social presence.

### Assessment of cybersickness

A small subgroup of VR users experiences cybersickness, such as headaches, nausea or disorientation [30]. The extent of adverse effects after exposure to a virtual environment has not been empirically investigated in the TBI population. The Simulator Sickness Questionnaire (SSQ) has been translated to Norwegian and will be used to assess cybersickness [67]. The questionnaire asks participants to score 16 symptoms on a four-point scale (0–3). SSQ will be administered before and after TASIT is administered for both 2D and VR versions and comparisons will be made to determine if the VR version has more adverse effects than the 2D version.

### Statistical analysis

Based on published data on the original version of TASIT [21, 33], there is reason to believe that the healthy control group scores will not be normally distributed, while TBI group scores will have a normal distribution. We will use paired sample t-tests in comparisons involving normally distributed continuous data and Mann–Whitney U tests when comparing skewed data.

*Construct validity* will be determined by known groups validity, convergent and divergent validity. *Known-groups validity* will be established by exploring differences between both VR and 2D TASIT and between patients with TBI and healthy controls using independent sample t-test or Mann–Whitney U test, depending on distribution of data. *Convergent validity* will be calculated as the correlation between VR TASIT results with established tests of emotion recognition and Theory of Mind, as well as self-reported empathy. *Divergent validity* will be calculated as the correlation between VR TASIT results with results on cognitive measures (processing speed, attention, working memory, abstract reasoning and executive functions) and measures of anxiety and depression symptoms.

*Test–retest reliability* will be calculated as the intraclass correlation coefficient between VR TASIT at T1 and T2.

*Ecological validity* will be calculated as correlation between VR TASIT results and self- and informant

reported results on measures of everyday social functioning and social communication.

*Presence* will be calculated as correlation between measures of self-reported levels of presence after exposure to VR TASIT and 2D TASIT.

*Cybersickness* will be calculated as correlation between measures of self-reported cybersickness after exposure to VR TASIT and 2D TASIT.

### Sample size and power calculation

Calculation of power using  $g^*$ power [68] has demonstrated that paired sample t-tests (e.g. test–retest in patients) would require a sample of 45 pairs, given a medium effect size,  $\alpha$ -value of 0.05, and power of 0.95. Given the planned group size of 50, we allow for an expected drop-out rate of 10% from T1 to T2. For the Mann–Whitney U tests, we have calculated power based on the group means reported by McDonald et al. [33], where controls had a mean score of 25 (SD 2), and patients had a mean of 19 (SD 5). Provided a medium effect size,  $\alpha$ -value 0.05 and power of 0.9, we would only need 9 patients to detect the same difference. However, we do not know that the Norwegian data will have the same score ratio, and this has never been done in VR, leaving a sample size of 50 in each group robust. As a strong relationship between VR and 2D TASIT is expected, we will pool data from VR and 2D TASIT in the correlational analysis (validity testing), giving a sample of 100 patients and 100 controls. This implies that we will be able to detect a weak correlation of  $r=0.25$  with a power of 0.9, given  $\alpha$ -value 0.05. In sub-analysis of VR and 2D TASIT separately ( $n=50$ ), a weak correlation could still be detected with a power of 0.08.

### Discussion

The purpose of this paper is to describe the development of a Norwegian VR test of social cognition, VR TASIT, and the protocol for the validation of VR TASIT in participants with TBI and in healthy controls. As the software has been successfully developed, the next step is to explore whether it has good construct validity, test–retest reliability and ecological validity. We will also explore the level of social presence experienced when exposed to VR TASIT and document the prevalence of adverse effects, i.e. cybersickness.

TASIT is one of few standardized tests of social cognition that recognizes the need for dynamic, multidimensional and contextually embedded assessment of social cognition in clinical populations at risk of impaired social cognition [32]. It is however limited by stimulus materials presented on a computer screen, a situation quite different from everyday social interaction. VR technology allows for a balance between the internal validity of

standardized test conditions and a naturalistic environment representative of everyday social behavior. It is hypothesized that the use of 360° videos with realistic social contexts in a head mounted display that eliminates distraction from outside stimuli increases the experience of social presence, and thus, ecological validity. In addition, there are practical benefits to computerized testing in general, in terms of automatization of administration, which provides clinicians more time for interpreting results and providing feedback and rehabilitation advice to patients.

Although dynamic and complex stimuli are more similar to everyday social situations than static pictures, and thus may also be more sensitive to everyday social cognitive impairment, it might well be the case that complex tasks at the same time introduces more noise to the measurement. For example, impaired attention, working memory, processing speed, and other cognitive functions may affect performance in addition to social cognitive problems. Thus, there is a possibility that more dynamic and complex tests may be less specific than more tests with higher levels of experimental control. A study that compared three emotion recognition tests in healthy people using static photographs, morphed photographs and videos as stimuli found only moderate correlations between the total scores of the three tests, suggesting that these stimuli might tap into different aspects of the emotion recognition construct [69].

To date, VR technology in healthcare has primarily been applied to medical training and treatment of conditions such as pain and anxiety [70]. VR is weakly established in neurorehabilitation, and although some VR interventions exist, they are characterized by few participants and lack of control groups [71]. The present study aims to implement VR in neurological rehabilitation using a systematic methodological research design, as well as systematically measuring any ill effects of VR exposure, both of which have been lacking in research on VR technology in health care in general [72].

The development of VR TASIT has benefited from collaboration between experts in brain injury rehabilitation, computer programming and film production. Further work remains before VR TASIT can be clinically implemented. The test's usability (i.e. user-friendliness) for patients and clinicians is yet to be systematically assessed. Both clinical practice and preliminary research indicates that persons with TBI in the chronic phase tolerate VR use well [31], but this remains to be investigated with regards to VR TASIT. In addition, in its current form the full test is lengthy, with an administration time of approximately 1 ½ hours. Thus, the total number of items may need to be reduced, which requires a systematic analysis to

determine which items can be eliminated without sacrificing validity. Furthermore, our overall aim of this work does not include establishment of normative data, which will ultimately be important for guiding clinicians in determining whether a patient has impaired social cognitive functioning. It is also an empirical question whether VR TASIT is sensitive to change in social cognition. This important question should be explored in future studies once the test is made available and has been validated. In summary, the development and validation of VR TASIT will be an important first step towards establishing a valuable clinical tool for assessment of social cognition. Finally, the relatively low costs of development of realistic everyday stimulus material indicates that this approach is of potential relevance to related research areas, both basic and applied.

#### Abbreviations

BRIEF-A	Behavioral Rating Inventory of Executive Functions-Adult version
CPT III	Conners' Continuous Performance Test, 3rd edition
CT	Computerized tomography
D-KEFS	Delis-Kaplan Executive Functions System
EET	The Emotional Evaluation Test
ERT	Emotion Recognition Task
GAD-7	General Anxiety Disorder 7
IRI	Interpersonal Reactivity Index
LCQ	LaTrobe Communication Questionnaire
MPS	Multimodal Presence Scale
MRI	Magnetic Resonance Imaging
PHQ-9	Patient Health Questionnaire
SIT-e	The Situational Inference Test-Enriched
SIT-m	The Situational Inference Test-Minimal
SRH	Sunnaas Rehabilitation Hospital
SSQ	Simulator Sickness Questionnaire
SSQ-TBI	Social Skills Questionnaire for Traumatic Brain Injury
TASIT	The Awareness of Social Inference Test
TBI	Traumatic Brain Injury
ToM	Theory of Mind
TSD	Tjenester for Sensitive Data (Service for Sensitive Data)
VR	Virtual Reality
WAIS IV	Wechsler Adult Intelligence Scale, 4th edition

#### Acknowledgements

We want to thank director and producer Tor Segelcke from StoriesToTell production company for producing all TASIT videos, as well as the crew and actors who contributed to the production. We also want to thank programmer Ivan Norderhaug for his contribution with software development, Sunnaasstiftelsen for funding of the VR video production, as well as clinical and administrative staff at Sunnaas Rehabilitation Hospital.

#### Authors' contributions

Author M.M. contributed to the conceptualization and design of the study, was responsible for software development and production of stimulus materials, and wrote the original draft and final revision of the manuscript. Authors M.L. and A.O. contributed to the original draft, as well as review and revisions of the manuscript. M.L. also contributed to software development and stimulus material production. Authors T.J., S.T. and S.M. contributed to the conceptualization and design of the study, stimulus production, as well as review and revision of the manuscript. Authors A.C.M., A.L., F.B., J.S., J.P. and D.N. reviewed and contributed to revision of the manuscript. C.B. has acted as supervising statistician in the design phase. All authors have read and approved the final version of the manuscript for submission.

## Funding

The corresponding author was funded by Stiftelsen Dam, grant FO387041. The development of VR TASIT was funded by Sunnaasstiftelsen.

## Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. The software and videos described in the article are available on request.

## Declarations

### Ethics approval and consent to participate

The study has been approved by the Regional Ethical committee, region South East (REC South East), Norway, REC number: 376999 and by Sikt—Norwegian Agency for Shared Services in Education and Research, previously NSD, number 172224. Informed consent will be obtained from all participants involved in the study. The study will be performed in accordance with relevant guidelines and regulations.

### Consent for publication

Written informed consent was obtained from individuals who are identifiable in the publication.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Research, Sunnaas Rehabilitation Hospital, Nesodden, Norway. <sup>2</sup>Department of Psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway. <sup>3</sup>Department of Occupational Therapy, Faculty of Health Sciences, Institute of Rehabilitation Science and Health Technology, Oslo Metropolitan University, Oslo, Norway. <sup>4</sup>Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway. <sup>5</sup>Clinic of Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. <sup>6</sup>NorHEAD - Norwegian Centre for Headache Research, Trondheim, Norway. <sup>7</sup>Department of Life Sciences and Health, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway. <sup>8</sup>Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>9</sup>Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway. <sup>10</sup>Department of Neurology, Subdepartment of Neuropsychology, University of Groningen, University Medical Center, Groningen, The Netherlands. <sup>11</sup>School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Clayton, Australia. <sup>12</sup>Monash-Epworth Rehabilitation Research Centre, Epworth Healthcare, Richmond, Australia. <sup>13</sup>Department of Physical Medicine and Rehabilitation, Indiana University School of Medicine, Indianapolis, IN, US. <sup>14</sup>School of Psychology, University of New South Wales, Kensington, Australia.

Received: 29 June 2023 Accepted: 2 August 2023

Published online: 07 September 2023

## References

- Frith CD, Frith U. Social Cognition in Humans. *Curr Biol*. 2007;17(16):R724–32. <https://doi.org/10.1016/j.cub.2007.05.068>.
- Babbage DR, Yim J, Zupan B, Neumann D, Tomita MR, Willer B. Meta-analysis of facial affect recognition difficulties after traumatic brain injury. *Neuropsychology*. 2011;25(3):277–85. <https://doi.org/10.1037/a0021908>.
- Milders M. Relationship between social cognition and social behaviour following traumatic brain injury. *Brain Inj*. 2019;33(1):62–8. <https://doi.org/10.1080/02699052.2018.1531301>.
- Ponsford JL, Downing MG, Olver J, Ponsford M, Acher R, Carty M, Spitz G. Longitudinal Follow-Up of Patients with Traumatic Brain Injury: Outcome at Two, Five, and Ten Years Post-Injury. *J Neurotrauma*. 2014;31(1):64–77. <https://doi.org/10.1089/neu.2013.2997>.
- Kelly M, McDonald S, Frith MH. Assessment and Rehabilitation of Social Cognition Impairment after Brain Injury: Surveying Practices of Clinicians. *Brain Impairment*. 2017;18(1):11–35. <https://doi.org/10.1017/Brlmp.2016.34>.
- Allain P, Togher L, Azouvi P. Social cognition and traumatic brain injury: Current knowledge. *Brain Inj*. 2019;33(1):1–3. <https://doi.org/10.1080/02699052.2018.1533143>.
- Henry JD, Cowan DG, Lee T, Sachdev PS. Recent trends in testing social cognition. *Curr Opin Psychiatry*. 2015;28(2):133–40. <https://doi.org/10.1097/YCO.0000000000000139>.
- McDonald S. Impairments in social cognition following severe traumatic brain injury. *J Int Neuropsychol Soc*. 2013;19(3):231–46. <https://doi.org/10.1017/s1355617712001506>.
- Osborne-Crowley K. Social cognition in the real world: Reconnecting the study of social cognition with social reality. *Rev Gen Psychol*. 2020;24:144–58. <https://doi.org/10.1177/1089268020906483>.
- Adolphs R. Conceptual Challenges and Directions for Social Neuroscience. *Neuron*. 2010;65(6):752–67. <https://doi.org/10.1016/j.neuron.2010.03.006>.
- Sohlberg MM, MacDonald S, Byom L, Iwashita H, Lemongello R, Meulenbroek P, Ness B, O'Neil-Pirozzi TM. Social communication following traumatic brain injury part I: State-of-the-art review of assessment tools. *Int J Speech Lang Pathol*. 2019;21(2):115–27. <https://doi.org/10.1080/17549507.2019.1583280>.
- Wallis K, Kelly M, McRae SE, McDonald S, Campbell LE. Domains and measures of social cognition in acquired brain injury: A scoping review. *Neuropsychol Rehabil*. 2022;32(9):2429–63. <https://doi.org/10.1080/09602011.2021.1933087>.
- Ekman P, Friesen WV. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*. 1976;1(1):56–75. <https://doi.org/10.1007/BF01115465>.
- Young A, Perrett D, Calder A, Sprengelmeyer R, Ekman P. Facial expressions of emotion: Stimuli and tests (FEEST). Bury St Edmunds: Thames Valley Test Company; 2002.
- Martin-Rodriguez JF, Leon-Carrion J. Theory of mind deficits in patients with acquired brain injury: A quantitative review. *Neuropsychologia*. 2010;48(5):1181–91. <https://doi.org/10.1016/j.neuropsychologia.2010.02.009>.
- Milders M, Fuchs S, Crawford JR. Neuropsychological impairments and changes in emotional and social behaviour following severe traumatic brain injury. *J Clin Exp Neuropsychol*. 2003;25(2):157–72. <https://doi.org/10.1076/jcen.25.2.157.1364212754675>.
- Milders M, Ietswaart M, Crawford JR, Currie D. Social behavior following traumatic brain injury and its association with emotion recognition, understanding of intentions, and cognitive flexibility. *J Int Neuropsychol Soc*. 2008;14(2):318–26. <https://doi.org/10.1017/S1355617708080351>.
- Kelly G, Brown S, Todd J, Kremer P. Challenging behaviour profiles of people with acquired brain injury living in community settings. *Brain Inj*. 2008;22(6):457–70. <https://doi.org/10.1080/02699050802060647>.
- Zaki J, Ochsner K. The Need for a Cognitive Neuroscience of Naturalistic Social Cognition. *Ann N Y Acad Sci*. 2009;1167:16–30. <https://doi.org/10.1111/j.1749-6632.2009.04601.x>.
- Barrett LF, Mesquita B, Gendron M. Context in Emotion Perception. *Curr Dir Psychol Sci*. 2011;20(5):286–90. <https://doi.org/10.1177/0963721411422522>.
- McDonald S, Flanagan S, Rollins J, Kinch J. TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *J Head Trauma Rehabil*. 2003;18(3):219–38. <https://doi.org/10.1097/00001199-200305000-00001>.
- Watts AJ, Douglas JM. Interpreting facial expression and communication competence following severe traumatic brain injury. *Aphasiology*. 2006;20(8):707–22. <https://doi.org/10.1080/02687030500489953>.
- Turner P. Affect, Availability and Presence. In M Lombard, F Biocca, J Freeman, W IJsselstein, RJ Schaevitz (Eds.), *Immersed in Media: Telepresence Theory, Measurement & Technology*. 2015;59–71. Springer International Publishing. [https://doi.org/10.1007/978-3-319-10190-3\\_4](https://doi.org/10.1007/978-3-319-10190-3_4).
- Ryan WS, Cornick J, Blascovich J, Bailenson JN. Virtual reality: whence, how and what for. In a. "Skip" Rizzo & S. Boucharde (Eds.), *Virtual Reality for Psychological and Neurocognitive Interventions*. 2019;15–46. Springer. [https://doi.org/10.1007/978-1-4939-9482-3\\_2](https://doi.org/10.1007/978-1-4939-9482-3_2).
- Riva G, Wiederhold BK, Mantovani F. Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine. *Cyberpsychol Behav Soc Netw*. 2019;22(1):82–96. <https://doi.org/10.1089/cyber.2017.29099.gri>.

26. Lombard M, Ditton T. At the heart of it all: the concept of presence. *J Comput-Mediated Commun.* 1997;3(2):0–0. <https://doi.org/10.1111/j.1083-6101.1997.tb00072.x>.
27. Biocca F, Harms C, Burgoon JK. Toward a more Robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators Virtual Environ.* 2003;12(5):456–80. <https://doi.org/10.1162/105474603322761270>.
28. Oh CS, Bailenson JN, Welch GF. A systematic review of social presence: definition antecedents and implications. *Front Robot AI.* 2018;5:409295. <https://doi.org/10.3389/frobt.2018.00114>.
29. Diemer J, Alpers GW, Peperkorff HM, Shiban Y, Mühlberger A. The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Front Psychol.* 2015;6:26. <https://doi.org/10.3389/fpsyg.2015.00026>.
30. Weech S, Kenny S, Barnett-Cowan M. Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Front Psychol.* 2019;10:158. <https://doi.org/10.3389/fpsyg.2019.00158>.
31. Greenhalgh M, Fitzpatrick C, Rodabaugh T, Madrigal E, Timmerman M, Chung J, Ahuja D, Kennedy Q, Harris OA, Adamson MM. Assessment of task demand and usability of a virtual reality-based rehabilitation protocol for combat related traumatic brain injury from the perspective of veterans affairs healthcare providers: a pilot study. *Front Virtual Reality.* 2021;2. <https://doi.org/10.3389/frvir.2021.741578>
32. McDonald S. New Frontiers in Neuropsychological Assessment: Assessing Social Perception Using a Standardised Instrument, The Awareness of Social Inference Test. *Aust Psychol.* 2012;47(1):39–48. <https://doi.org/10.1111/j.1742-9544.2011.00054.x>.
33. McDonald S, Bornhofen C, Shum D, Long E, Saunders C, Neulinger K. Reliability and validity of The Awareness of Social Inference Test (TASIT): a clinical test of social perception. *Disabil Rehabil.* 2006;28(24):1529–42. <https://doi.org/10.1080/09638280600646185>.
34. Bliksted V, Videbech P, Fagerlund B, Frith C. The effect of positive symptoms on social cognition in first-episode schizophrenia is modified by the presence of negative symptoms. *Neuropsychology.* 2017;31(2):209–19. <https://doi.org/10.1037/neu0000309>.
35. Kipps CM, Nestor PJ, Acosta-Cabrero J, Arnold R, Hodges JR. Understanding social dysfunction in the behavioural variant of frontotemporal dementia: The role of emotion and sarcasm processing. *Brain.* 2009;132(3):592–603. <https://doi.org/10.1093/brain/awn314>.
36. McDonald S, Flanagan S. Social perception deficits after traumatic brain injury: Interaction between emotion recognition, mentalizing ability, and social communication. *Neuropsychology.* 2004;18(3):572–9. <https://doi.org/10.1037/0894-4105.18.3.572>.
37. McDonald S, Saunders JC. Differential impairment in recognition of emotion across different media in people with severe traumatic brain injury. *J Int Neuropsychol Soc.* 2005;11(4):392–9. <https://doi.org/10.1017/S1355617705050447>.
38. Risko EF, Richardson DC, Kingstone A. Breaking the Fourth Wall of Cognitive Science: Real-World Social Attention and the Dual Function of Gaze. *Curr Dir Psychol Sci.* 2016;25(1):70–4. <https://doi.org/10.1177/0963721415617806>.
39. Bormann D, Greitemeyer T. Immersed in Virtual Worlds and Minds: Effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. *Social Psychol Pers Sci.* 2015;6(6):646–52. <https://doi.org/10.1177/1948550615578177>.
40. Jung S, Lindeman RW. Perspective: does realism improve presence in vr? suggesting a model and metric for vr experience evaluation. *Front Virtual Real.* 2021;2:693327. <https://doi.org/10.3389/frvir.2021.693327>.
41. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc B: Biol Sci.* 2009;364(1535):3549–57. <https://doi.org/10.1098/rstb.2009.0138>.
42. Argyriou L, Economou D, Bouki V. Design methodology for 360° immersive video applications: the case study of a cultural heritage virtual tour. *Pers Ubiquit Comput.* 2020;24(6):843–59. <https://doi.org/10.1007/s00779-020-01373-8>.
43. Fonseca D, Kraus M. A comparison of head-mounted and hand-held displays for 360° videos with focus on attitude and behavior change. *Proceedings of the 20th International Academic Mindtrek Conference.* 2016 287–296. <https://doi.org/10.1145/2994310.2994334>
44. Silverberg ND, Iverson GL, Cogan A, Dams-O'Connor K, Delmonico R, Graf MJP, Iaccarino MA, Kajankova M, Kamins J, McCulloch KL, McKinney G, Nagele D, Panenka WJ, Rabinowitz AR, Reed N, Wethe JV, Whitehair V, Anderson V, Arciniegas DB, Zemek R. The american congress of rehabilitation medicine diagnostic criteria for mild traumatic brain injury. *Arch Phys Med Rehabil.* 2023. <https://doi.org/10.1016/j.apmr.2023.03.036>
45. University of Oslo. Short introduction to Nettskjema. 2023. <https://www.uio.no/english/services/it/adm-services/nettskjema/about-nettskjema.html>
46. Corcoran R, Mercer G, Frith CD. Schizophrenia, symptomatology and social inference: Investigating “theory of mind” in people with schizophrenia. *Schizophr Res.* 1995;17(1):5–13. [https://doi.org/10.1016/0920-9964\(95\)00024-G](https://doi.org/10.1016/0920-9964(95)00024-G).
47. Frøyhaug M, Andersson S, Andreassen OA, Ueland T, Vaskinn A. Theory of mind in schizophrenia and bipolar disorder: psychometric properties of the Norwegian version of the hinting task. *Cogn Neuropsychiatry.* 2019;24(6):454–69. <https://doi.org/10.1080/13546805.2019.1674645>.
48. Kessels RP, Montagne B, Hendriks AW, Perrett DI, de Haan EH. Assessment of perception of morphed facial expressions using the emotion recognition task: normative data from healthy participants aged 8–75. *J Neuropsychol.* 2014;8(1):75–93. <https://doi.org/10.1111/jnp.12009>.
49. Rosenberg H, Dethier M, Kessels RP, Westbrook RF, McDonald S. Emotion perception after moderate-severe traumatic brain injury: the valence effect and the role of working memory, processing speed, and nonverbal reasoning. *Neuropsychology.* 2015;29(4):509–21. <https://doi.org/10.1037/neu0000171>.
50. Davis MH. A multidimensional approach to individual differences in empathy. *JSAS Catal Sel Doc Psychol.* 1980;10:85.
51. Wechsler D. Wechsler adult intelligence scale—Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson. 2008;22(498):1.
52. Conners CK. Conners' continuous performance test – 3rd edition (CPT-3) manual. Toronto: Multi-Health Systems; 2014.
53. Delis DC, Kaplan E, Kramer JH. Delis-Kaplan Executive Function System: Technical Manual. San Antonio: Harcourt Assessment Company; 2001.
54. Roth RM, Isquith PK, Gioia GA. Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A). Lutz: Psychological Assessment Resources; 2005.
55. Bora E, Berk M. Theory of mind in major depressive disorder: a meta-analysis. *J Affect Disord.* 2016;191:49–55. <https://doi.org/10.1016/j.jad.2015.11.023>.
56. Plana I, Lavoie M-A, Battaglia M, Achim AM. A meta-analysis and scoping review of social cognition performance in social phobia, post-traumatic stress disorder and other anxiety disorders. *J Anxiety Disord.* 2014;28(2):169–77. <https://doi.org/10.1016/j.janxdis.2013.09.005>.
57. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092–7. <https://doi.org/10.1001/archinte.166.10.1092>.
58. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606–13. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
59. Ietswaart M, Milders M, Crawford JR, Currie D, Scott CL. Longitudinal aspects of emotion recognition in patients with traumatic brain injury. *Neuropsychologia.* 2008;46(1):148–59. <https://doi.org/10.1016/j.neuro-psychologia.2007.08.002>.
60. Hanks RA, Temkin N, Machamer J, Dikmen SS. Emotional and behavioral adjustment after traumatic brain injury. *Arch Phys Med Rehabil.* 1999;80(9):991–7. [https://doi.org/10.1016/s0003-9993\(99\)90049-7](https://doi.org/10.1016/s0003-9993(99)90049-7).
61. Katz MM, Lyerly SB. Methods for measuring adjustment and social behavior in the community: I. Rationale, description, discriminative validity and scale development. *Psychol Rep.* 1963;13:503–35. <https://doi.org/10.2466/pr0.1963.13.2.503>.
62. Francis HM, Osborne-Crowley K, McDonald S. Validity and reliability of a questionnaire to assess social skills in traumatic brain injury: a preliminary study. *Brain Inj.* 2017;31(3):336–43. <https://doi.org/10.1080/02699052.2016.1250954>.
63. Douglas JM, O'Flaherty CA, Snow PC. Measuring perception of communicative ability: the development and evaluation of the La Trobe communication questionnaire. *Aphasiology.* 2000;14(3):251–68. <https://doi.org/10.1080/026870300401469>.
64. Hansen SM, Stubberud J, Hjerstedt M, Kirmess M. Intensive and standard group-based treatment for persons with social communication difficulties after an acquired brain injury: study protocol for a randomised

- controlled trial. *BMJ Open*. 2019;9(9):e029392. <https://doi.org/10.1136/bmjopen-2019-029392>.
65. Douglas JM, Bracy CA, Snow PC. Measuring perceived communicative ability after traumatic brain injury: reliability and validity of the La trobe communication questionnaire. *J Head Trauma Rehabil*. 2007;22(1):31–8. <https://doi.org/10.1097/00001199-200701000-00004>.
  66. Makransky G, Lilleholt L, Aaby A. Development and validation of the multimodal presence scale for virtual reality environments: a confirmatory factor analysis and item response theory approach. *Comput Human Behav*. 2017;72:276–85. <https://doi.org/10.1016/j.chb.2017.02.066>.
  67. Brown P, Spronck P, Powell W. The simulator sickness questionnaire and the erroneous zero baseline assumption. *Front Virtual Real*. 2022;3:118. <https://doi.org/10.3389/frvir.2022.945800>.
  68. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical Power Analyses Using G\*Power 3.1: Tests for Correlation and Regression Analyses. *Behav Res Methods*. 2009;41:1149–60. <https://doi.org/10.3758/BRM.41.4.1149>.
  69. Khosdelazad S, Jorna LS, McDonald S, Rakers SE, Huitema RB, Buunk AM, Spikman JM. Comparing static and dynamic emotion recognition tests: Performance of healthy participants. *PLoS ONE*. 2020;15(10):e0241297.
  70. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, Glisic M, Zeiner S, Kulnik ST, Crutzen R, Kimberger O, Kletecka-Pulker M, Atanasov AG, Willschke H. Virtual and augmented reality applications in medicine: analysis of the scientific literature. *J Med Internet Res*. 2021;23(2):e25499. <https://doi.org/10.2196/25499>.
  71. Riva G, Mancuso V, Cavedoni S, Stramba-Badiale C. Virtual reality in neurorehabilitation: A review of its effects on multiple cognitive domains. *Expert Rev Med Devices*. 2020;17(10):1035–61. <https://doi.org/10.1080/17434440.2020.1825939>.
  72. Vlaker JH, van Bommel J, Riva G, Wiederhold BK, Cipresso P, Rizzo AS, Botella C, Hooft L, Bienvenu OJ, Geerts B, Wils E-J, Gommers D, van Genderen ME. Reporting the early stage clinical evaluation of virtual-reality-based intervention trials: RATE-VR. *Nat Med*. 2023;29(1):12–3. <https://doi.org/10.1038/s41591-022-02085-7>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

