Comprehensive Psychiatry 119 (2022) 152350

Contents lists available at ScienceDirect

Comprehensive Psychiatry

journal homepage: www.elsevier.com/locate/comppsych

Virtual reality-based theory of mind intervention in schizophrenia: Preliminary efficacy results

Edit Vass [a,*], Viktória Simon [a], Gábor Csukly [a], Zita Fekete [b], Balázs Kis [c], Lajos Simon [a]

[a] Semmelweis University, Faculty of Medicine, Department of Psychiatry and Psychotherapy, 1082 Balassa str 6, Budapest, Hungary
[b] University of Debrecen, Institute of Behavioral Sciences, 22 Móricz Zsigmond körút, Debrecen 4032, Hungary
[c] Szabolcs-Szatmar-Bereg County Hospital and University Teaching Hospital, Department of Psychiatry, 68 Szent István street, Nyíregyháza 4400, Hungary

ARTICLE INFO

Keywords:
Theory of Mind
schizophrenia
virtual reality

ABSTRACT

Background: In recent years, a growing body of literature has supported the core nature and functional significance of Theory of Mind (ToM) deficit in schizophrenia. These findings have made ToM impairment a promising treatment target. However, despite the encouraging results, its complexity makes it difficult to develop new interventions and even to understand the exact nature and scope of the deficit. Yet, further investigation has suggested that using modern technology and multilevel assessment may help solve the problem.
Methods: Virtual Reality-based Theory of Mind Intervention (VR-ToMIS) is a recently developed structured method using the combination of cognitive and behavioral therapeutic techniques and the advantages of Virtual Reality (VR) technology. A controlled study with a three-month follow-up was conducted with 42 patients (suffering from schizophrenia or schizo-affective disorder) randomly assigned to either an experimental (VR-ToMIS) or control group (passive-VR). Repeated two-way factorial analysis of covariance was used to evaluate the effects of VR-ToMIS on symptoms, neuro- and social cognition, pragmatic skills, and quality of life when the effect of IQ was controlled.
Results: Patients participating in VR-ToMIS showed significant improvements in all types of ToM tasks (except for hyper-ToM task, based on the results of Cartoon test, Faux pas test and and Baron-Cohen Minds in the Eyes Test) compared to the control group with moderate to large effect sizes. In the case of negative and cognitive symptoms, significant between-group differences were also supported. Improvement was moderated by IQ in the case of higher-order ToM, manner, and relevance implicatures. Results were proved to be sustainable three months after the treatment.
Conclusion: Although the presented results are considered preliminary, they support the potential of the integration of modern technology and traditional methods for future interventions.

1. Introduction

Schizophrenia is a severe mental disorder usually characterized by recurrent psychotic episodes. It affects up to 1% of the population and is associated with a variety of social-cognitive, neurocognitive, and functional deficits, often leading to disability, and representing a significant societal burden. [1]

Over recent decades, researchers have turned their attention to interventions focusing on social cognitive deficits, as this may be an important element of the patients' recovery process in schizophrenia [2,3]. Among the fields of social cognition, Theory of Mind (ToM) appears to be particularly important, as it has been identified as a significant predictor of social functioning [2,4,5].

ToM represents a complex cognitive domain that pertains to our ability to infer and understand our own mental states as well as others', through playing a key role in the interpretation and prediction of others' behavior [6]. Several studies have aimed to explore the scope and nature of ToM deficits in schizophrenia [7–9]. These researches indicate that

Abbreviations: VR, Virtual reality; VR-ToMIS, Virtual Reality-Based Theory of Mind Intervention in Schizophrenia; RMET, Baron-Cohen Minds in the Eyes Test; M.I.N·I., MINI International Neuropsychiatric Interview; WCST-64, Wisconsin Card Sorting Test; RBANS, Repeated Battery for the Assessment of Neuropsychological Status; LQoLP, Lancashire Quality of Life Profile; rANCOVA, Repeated measures of analysis of covariance; SSQ, Simulator Sickness Questionnaire,; TDC, Temporal Disc Controller.

* Corresponding author at: 1082 Balassa str 6, Budapest, Hungary.
E-mail address: vass.edit@med.semmelweis-univ.hu (E. Vass).

https://doi.org/10.1016/j.comppsych.2022.152350

Available online 18 October 2022

ToM is not a homogeneous construct, but rather an integration of different processes. This recognition has led to the emergence of concepts like cognitive- and affective-ToM, differentiating ToM aspects along with the characteristics of the inferred mental states, where the separability of the two processes is also supported by neuroimaging studies [10,11]. ToM can be classified hierarchically as well, along with the construct of intentionality [8,12]. In this respect, differentiation can be made between lower- and higher-order ToM, e.g.: first-, second-, or third-order ToM, where first-order processes indicate one's knowledge of someone as having a mental state, second-order processes refer to one's understanding that someone's mental state might be different from one's own, and so on [4,13]. Furthermore, Abu-Akel et al. (2013) offered an additional approach to scope the complexity of ToM, suggesting an extension of its assessment to the levels of the process, including representation, attribution, and application of mental states [7]. Despite the expansion of knowledge about ToM, no measuring tool has been developed that covers all of the mentioned aspects [3]. However, the diversity of the ToM tasks used in different studies may reflect the complexity and heterogeneity of ToM. This may also explain the conflicting efficacy outcomes from different studies on different interventions. However, the limited scope of the assessment tools might not be the only explanation, as grasping the complexity of ToM can also be challenging for the developers of the interventions.

The development of methods to improve ToM follows two different approaches. The so-called broad-based interventions aim to cover all neuro- and/or socio-cognitive fields associated with functional deficits. In contrast, targeted interventions focusing on one area attempt to influence functionality directly through the remediation of it. Both approaches have limitations. Comprehensive approaches are less able to reflect on the individual patterns of cognitive impairment. On the other hand, although targeted interventions are used in patients with a particular deficiency, this approach can limit the range of patients for whom they can be used effectively. Hence, a combination of interventions with narrowed focuses that consider individual differences might be a promising alternative, highlighting the importance of the development of more effective targeted methods [3,14].

Even if early studies support the possibility that ToM impairment is remediable, only a few interventions focus selectively on ToM skills [15–19]. Based on our previous review, compared to methods that follow a broad-based approach, targeted interventions led to more stable results in terms of their impact on functionality and ToM, even if they used different types of assessment tools [3]. Such interventions typically use short videotapes or comic strip tasks to illustrate social situations, where patients are asked to interpret the presented situations with the assistance of the therapist. Although results suggest that these ToM-focused methods might be more able to capture its heterogeneous nature, in the case of higher-order ToM skills, mixed results were reported, indicating that even these methods are only partially able to capture the complexity of ToM [15,20,21]. As some authors suggest, the efficacy of novel interventions may be further increased by changing some factors. These changes might involve making the simulations more realistic, using verbalization to help the patient extract the relevant information from the characters' mental state, or enabling the analysis of social interactions from a first-person perspective [3,15,18].

With these factors in mind, our research team undertook to develop a VR-based targeted ToM intervention based on realistic simulations of social interactions, allowing a first-person experience and taking the strengths and weaknesses of the previously developed methods into account. We aimed to develop a complex intervention that is able to capture the complex nature of ToM [22]. To equally reflect the complexity of the assessment, four different ToM tools were used. We hypothesized that the use of VR-ToMIS would be associated with sustainable improvement in ToM, pragmatic language skills, negative symptoms and functional outcomes. Considering the contradictions of the literature data, we didn't assume a generalization effect on neurocognition. [3]

2. Methods

A randomized, controlled, single-blind clinical trial was performed.

Patients were recruited from the Department of Psychiatry and Psychotherapy, Semmelweis University, Budapest, and from the András Jósa University Teaching Hospital, Nyíregyháza, from March 2019 to October 2021. Participants were under treatment at the outpatient care departments of the mentioned hospitals at baseline and during the entire duration of the study. Patients were recruited on the recommendation of the psychiatrist treating them. Before participation, all subjects provided written informed consent to the study protocol reviewed and approved by the local Ethics Committee (ref. SE-TUKEB 150/2016. Date of approval: 3. Aug. 2016.).

To ensure balanced sampling, blocked randomization was used. Randomization was carried out by an independent psychiatrist of the András Jósa Teaching Hospital, using a random number generator-based Python software written for this study.

All participating patients met DSM 5 criteria for schizophrenia or schizoaffective disorder based on the patient's medical records and MINI International Neuropsychiatric Interview (M.I.N.I). Additional inclusion criteria were: (1) No documented hospitalization or change in their medication in the last 3 months, (2) Adequate language skills, (3) ToM deficit, identifiable either by Baron-Cohen Minds in the Eyes Test (RMET, scores $\leq$22 [23]), or by Faux pas test (correct answer rate $\leq$ 0.75 on any subscale [24]), and (4) Age 18–65 years. Exclusion criteria included: (1) Noncompliant behavior, (2) History of any neurological disease that might affect cognitive performance, (3) Diagnosis of epilepsy, (4) Comorbid alcohol/drug dependence/abuse, (5) mental retardation, and (6) participation in other clinical trials over the last 6 months.

2.1. Assessment

After the attending psychiatrist referred the patient to the research team, and the informed consent form was signed by the patient, each assessment was carried out by a clinical psychologist or a psychiatrist kept blinded to the allocation. Assessments took place at the Outpatient Care Unit of Semmelweis University or the András Jósa Teaching Hospital. The only exception was the Simulator Sickness Questionnaire, which was recorded right after each session by the psychotherapist providing the intervention.

2.1.1. Symptoms

The severity of the symptoms was assessed by using the Positive and Negative Syndrome Scale (PANSS), where according to the suggestion of Shafer and Dazzi (2019), positive, negative, cognitive, affective, and activity/excitement components were calculated [25]. The scale was administered by a trained clinician based on a clinical interview.

2.1.2. Neurocognitive assessment

IQ was measured at baseline by using the Hungarian version of the Wechsler Adult Intelligence Scale. The scale is administered by a trained clinician and examines the following domains: general knowledge, problem-solving, immediate memory, calculation skills, processing speed, recognition of causal relationships, and analytical skills. (Internal consistency: Cronbach's $\alpha = 0.94$, good reliability for the following constructs: immediate memory, processing speed, reasoning (α = 0.76–0.090)) [26].

A sequence of two neurocognitive tests was also administered: (1) Repeatable Battery for the Assessment of Neuropsychological Status (RBANS, internal consistency: $r = 0.81$, inter-rater reliability: $r = 0.68$), which provides age-adjusted scores for the measure of immediate and delayed memory, visuospatial skills, language skills, and attention, and (2) Wisconsin Card Sorting Test (WCST-64), which is used to assess abstract reasoning and perseveration. During the WCST-64 test, the patient - without prior knowledge of the rule - is asked to sort a series of

cards. In determining the rule, the patient can only rely on the clinician's feedback. (Good construct validity, Inter-rater reliability: r = 0.57–0.90). [27–31]

2.1.3. ToM assessment

To cover all possible aspects of ToM, multiple tests were used:

(1) Baron-Cohen Minds in the Eyes Test (RMET), which focuses on visual stimuli-based emotional state attribution (internal consistency $\alpha = 0.75$, inter-rater reliability r:0.80). During the test 36 eye regions of different people are presented to the patient, accompanied by four words expressing an emotional state. The patient is then asked to choose the word best suited to the displayed image. [32]
(2) Faux pas test, which is a complex test including components to assess the ability to recognize and understand social faux pas situations, and the ability to infer intentions and emotional states of another person. (Internal consistency: $\alpha = 0.94$ for faux pas stories, $\alpha = 0.92$ for non-faux pas stories, inter-rater reliability: $r = 0.88$ for faux pas stories, $r = 0.96$ for non-faux pas stories.) 20 short stories (10 of them contains a faux pas situation) are presented to the patient, who is asked to answer a series of questions about the intentions and emotional states of the characters of the stories. [33,34]
(3) Finally, the Cartoon stories task, which includes components to evaluate different levels of ToM [23,35]. (internal consistency: $\alpha = 0.80$, inter-rater reliability: $r = 0.63$) Cartoon test is a picture sequencing task followed by questions about the situation and the mental states of the characters with the aim of testing the patient's ability to infer mental processes.

2.1.4. Assessment of communicational pragmatic skills

To rule out the possible overlap between ToM and pragmatic assessment, the Hungarian Metaphor, and Irony test was used to assess pure pragmatic skills. Although the test itself contains ToM tasks as well, it focuses mainly on detecting and interpreting violations of Gricean maxima (implicatures). As its name suggests, the test also includes tasks that assess metaphor and irony comprehension [36]. There is conflicting literature that disputes whether these types of tasks should be used as a ToM or as a pragmatic assessment tool. However, as Bosco (2018) points out, understanding irony and metaphor require an understanding of someone else's point of view, indicating that these tools measure both constructs. [37] Following this line of reasoning, we decided to use these test components as part of the ToM evaluation, while the detection and interpretation of implicatures were considered as pure pragmatic assessment tools. During the test, short stories are presented to the patient, followed by a series of questions to test the ability to comprehend irony, metaphor, and supposed meaning.

2.1.5. Quality of Life (QoL) assessment

Lancashire Quality of Life Profile (LQoLP) was used to assess the level of a patient's subjective satisfaction in eleven fields, including job, safety, finances, health, or living conditions. Cantrel's ladder is also included in the test to assess overall satisfaction on a visual analog scale. The test is self-administered, where the patients are asked to indicate satisfaction with each examined field on a 7-point Likert scale. (internal consistency: $\alpha = 0.93$, inter-rater reliability: $r = 0.92$). [38,39]

For further, objective evaluation, patients' relatives, co-workers, or friends (chosen by the patient) were asked to assess the change after the intervention on a 5-point Likert scale on various dimensions, focusing primarily on self-determination and communicational skills. To ensure a freer expression of opinion, open-ended questions were also used. The questionnaire was designed by our research team, with the development goals of the intervention in mind.

All assessment tools were administered both at the baseline and after the intervention, except for IQ and objective QoL assessment, where the IQ test was administered only at the baseline, and the objective QoL test only after the intervention.

2.1.6. Assessing possible side effects

To monitor the possible side-effects, a Simulator Sickness Questionnaire (SSQ) was used, where each symptom was rated from 0 to 3, where 0 indicated the lack of symptoms, and scores from 1 to 3 referred to the severity of the symptoms (internal consistency: $\alpha = 0.86$–0.94, Inter-rater reliability r: 0.31–0.45) [40,41].

At the end of all sessions, patients were asked about their opinion on the method. Similar to the case of QoL, for the purpose of evaluation 7-point Likert scales and open-ended questions were used.

2.2. Intervention

Interventions were provided by trained psychotherapists.

VR-ToMIS is a 9-session long, structured intervention provided in 1-h individual sessions per week (immersive virtual environments were provided by vTime (available at: vtime.net)). The method consists of 1 preparative session and 8 virtual sessions (developed by using iterative testing process). Virtual sessions are based on three main steps. First, each patient participates in a virtual roleplay simulating everyday social interactions with an avatar controlled by the therapist. However, the therapist only guided the avatar's movements, and timed the avatar's verbal responses. Avatar's reactions were designed to be simple and to be capable of triggering the appearance of ToM and pragmatic impairment. To achieve this, dialogs included - for example - irony, metaphors, and double-meaning sentences [22]. Although vTime was aware of the development and we have their support, complete automation of the avatars could not be solved. Hence we used pre-recorded audios to voice the avatar. At this step, Samsung Gear Head Mounted Displays were used with a Samsung S7 or S8 smartphone and a Samsung Simple Controller. After the roleplay, patients were asked to visualize the emotional state inferred to the avatar, by using interactive software, called Temporal Disc Controller (TDC). During the task, the participant sees a 3D face on a screen. The facial expression of the 3D face can be changed by moving the pointing tool over the image, and at the same time over a circular interface underlying it. Moving the pointing tool in the center of the interface displays a neutral emotion, and moving towards the periphery, a wide range of mixtures of basic emotions with different intensities can be visualized [42]. After this task, experiences of the previous two steps and possible ways to overcome the difficulties are discussed with the therapist. In this phase, mainly cognitive and metacognitive techniques are used. When all three steps are finished, the patient has the opportunity to try and practice alternative reactions by repeating the previously mentioned steps [22].

In the control group, patients use the same virtual software. However, their activity is limited to the exploration of the virtual environment. It means that they can build their avatars and change between destinations, but they cannot contact other users via their avatars, and the therapist does not provide any intervention other than technical assistance.

Patients were blinded to the allocation.

3. Statistical analyses

3.1. Sample size calculation

Considering that the literature has no VR based ToM intervention or training similar to VR-ToMIS, the sample size was calculated from the pilot study with an α set at 0.05, a β set at 0.20 (that would provide an 80% chance of showing a statistically significant result) and an effect size of 0.84 (Cohen's d) for the overall under mentalizing score of Faux pas. [43] Using the method suggested by Shieh for sample size calculation for ANCOVA, the proposed study required a total sample size of 36 participants. With an estimated drop out rate of 15%, 42 participants

were estimated to be needed. [44]

3.2. Data analysis

Initial differences in main demographic and clinical characteristics between active and control conditions were evaluated by using either student *t*-test or Wilcoxon rank-sum test.

The effect of the intervention on symptoms, neuro- and social cognitive skills, and functional outcome was tested using repeated measures of analysis of covariance (two-way rANCOVA) procedure, where the post-treatment and follow-up scores of the outcome variables were included as the dependent variable, treatment condition as a between-group variable, and baseline scores of outcome variables as a continuous covariate [45]. Since the potential mediator role of IQ in ToM remediation is frequently discussed in the literature, to test its potential role, IQ and two-way interaction (treatment condition by moderator) were also entered as covariates. As a correction for multiple comparisons, the Bonferroni method was used [46].

To characterize between-group differences, partial eta squared was calculated, where $\eta_p^2 \leq 0.01$ represents small, $\eta_p^2 = 0.06$–0.14 is considered a medium, and $\eta_p^2 \geq 0.14$ indicates a large effect size. For further description of the effect of the intervention, Cohen's d was calculated in both active and control conditions, where $d \leq 0.2$ represents small, $d \geq 0.5$ indicates a medium, and ≥ 0.8 is considered a large effect size [47].

All statistics were performed by SAS, version: 9.04.01.

4. Results

Out of the 102 patients screened, 43 were included in the study. Participating patients were randomly allocated to the active VR-ToMIS group ($n = 22$) or the passive- VR group ($n = 21$). One patient dropped out from the active group due to an adverse event, not related to the intervention. (Trauma-related psychotic regression) Hence, the final sample included 21 patients in each group (Fig. 1). Two further patients were lost to follow-up, both in the control group.

According to a baseline comparison, no significant differences were found between the two groups. 22 male and 20 female patients aged between 18 and 63 participated in the study, where the mean illness duration was 20.98 years (within the range of 18–47 years). The mean score of the PANSS scale was 52.04 points, indicating mild symptoms (Table 1). The mean dose of antipsychotic medication expressed in chlorpromazine equivalent was 521.88 mg/day, which corresponds to a medium dose, where 83.3% of the patients took atypical, 2.38% typical antipsychotics and 14.28% took the combination of the two types of antipsychotic medication [47]. No change in medication was reported during the intervention in either group.

4.1. Efficacy of VR-ToMIS

4.1.1. Symptoms

As can be seen in Table 1, the sample was mainly characterized by negative and cognitive symptoms. Although the predominance of these types of symptoms didn't change after the intervention, ANCOVA indicated significant between-group differences with large effect sizes in both symptom groups in favor of VR-ToMIS (negative symptoms $F(1,41) = 8.67$, $p = .005$, $\eta_p^2 = 0.19$, cognitive symptoms $F(1,41) = 9.67$, $p = .003$, $\eta_p^2 = 0.21$) In the case of positive or affective symptoms, no significant change was detected.

4.1.2. Neurocognition

Immediate memory scores of the RBANS test ($F(1.41) = 5.99$, p:0.01, $\eta_p^2 = 0.16$) showed significant improvement in the VR-ToMIS condition, with a large effect size, where the significant moderator role of IQ wasn't confirmed. Scores of other neurocognitive skills didn't change significantly.

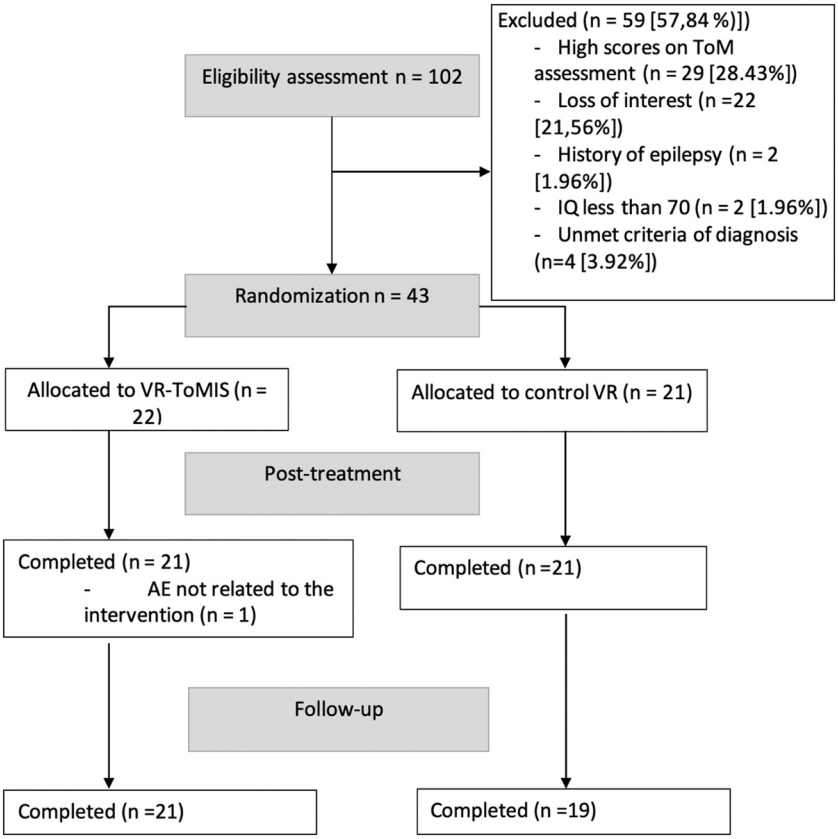

Fig. 1. Study flow-chart.

Table 1
Main demographic and clinical variables at baseline.

	VR-ToMIS (n:21)	Passive-VR (n:21)	Active/Control	
			Statistics (Fisher exact test, two sample *t*-test or Wilcoxon Rank Sum test)	p
Male/Female	12/9	10/11	OR: 0.20	0.76
Duration	17.31 (11.19)	19.76 (10.29)	t: −0.74	0.46
Age	36.71 (11.73)	42.47 (8.74)	t: −1.80	0.08
Education	14.09 (2.46)	15.09 (4.74)	z: −0.24	0.81
IQ	106 (10.03)	97.57 (14.21)	z: 1.91	0.06
Baron-Cohen Eyes Test	19.14 (4.15)	18.61 (3.29)	t: 0.45	0.65
Faux pas overall scores				
Hypo-ToM	0.53 (0.24)	0.47 (0.21)	t: 0.79	0.44
Hyper-ToM	0.93 (0.12)	0.89 (0.16)	z: 1.21	0.22
Cartoon test overall scores	41.95 (10.07)	40.42 (10.29)	t: 0.48	0.63
First-order ToM	3.81 (1.25)	3.42 (1.25)	z: 1.06	0.29
Second order ToM	3.66 (0.85)	3.57 (1.08)	z: 0.16	0.87
Third order ToM	1.14 (0.96)	1.24 (1.14)	z: −0.17	0.86
Metaphor-irony test				
Metaphor comprehension	7.38 (1.85)	6(2.84)	t: 1.86	0.07
Irony comprehension	8.28 (2.33)	6.81 (3.51)	z: 1.34	0.17
ToM sub-score	18.72 (7.21)	17.61 (5.38)	t: 0.41	0.68
PANSS (Shafer et al. 2009)	52.05 (10.19)	51.81 (9.87)	z: 0.06	0.94
Negative symptoms	15.28 (4.96)	15.33 (3.61)	t: −0.04	0.97
Positive symptoms	8.9(2.47)	8.9(2.47)	t: 0.0	1.0
Cognitive symptoms	13.61 (3.27)	14.57 (3.07)	t: −0.97	0.34
Activity/ Excitement	4.85 (1.01)	5.57 (0.31)	z: −1.83	0.07
Affective symptoms	9.19 (2.63)	8.33 (3.61)	z: 1.19	0.23

4.1.3. Theory of mind

Regardless of the type of assessment and the measured aspect of ToM, significant between-group differences were supported by ANCOVA in all cases (RMET, Faux pas subscales, ToM subscales of the Cartoon test, and the Metaphor and Irony test), with medium-to-large between- and within-group effect sizes ($\eta_p^2 = 0.12–0.49$, d = 0.57–1.17) (Table 2) There was only one exception. Faux pas task, which is associated with hyper-mentalization, remained unchanged in both groups. ($F(1,41) = 0.01$, $p = .93$). The significant moderator role of IQ was supported in the case of higher-order ToM tasks. (Metaphor comprehension: $F(1,41) = 23.89$, $p < .0001$, Irony comprehension: $F(1.41) = 9.30$, $p = .003$, ToM score, related to the violence of implicatures: $F(1,41) = 5.59$, $p = .02$, Third-order ToM task of Cartoon test: $F(1,41) = 8.43$, $p = .006$, Faux pas understanding: $F(1,41) = 4.86$, $p = .03$).

4.1.4. Pragmatic skills

In the case of all four measured types of implicatures measuring significant within and between-group differences were supported, with a large effect size. (Quantity implicature: $F(1,41) = 15.46$, $p = .0004$, $\eta_p^2 = 0.30$, Quality implicature: $F(1,41) = 24.22$, $p < .0001$, $\eta_p^2 = 0.58$, Manner implicature: $F(1,41) = 44.80$, $p < .0001$, $\eta_p^2 = 0.56$, Relevance implicature: $F(1,41) = 24.17$, $p < .0001$, $\eta_p^2 = 0.40$) Furthermore, in the case of two types of implicatures, improvement in the active group was accompanied by a significant decrease in the control group (Relevance implicature: $t = -3.36$, $p = 003$, quality implicature: t:-2.49, $p = 02$) Moderator role of IQ was also found in the case of manner ($F(1,41) = 4.76$, $p = .03$) and relevance implicature ($F(1,41) = 12.00$, $p = .001$).

4.1.5. Quality of Life

Considering subjective satisfaction with certain areas of life, no significant between- or within-group differences were shown (Table 3).

Twelve relatives/friends/colleagues of the patients responded to the questionnaire on the perceived change after the intervention. Based on questions about conversational skills, 75%–83.40% of the respondents detected at least some improvement. Concerning the degree of suspicion and misunderstanding, 41.60%–66.70% of the interviewees found that the patient tended to better understand others' mental states. The third group of questions was asked about the degree of independence. In this case, 41.70%–66.70% of the respondents answered that improvement was detectable (Table 4).

4.1.6. Simulator sickness and subjective feedback on VR-ToMIS

Only mild symptoms of simulator sickness were reported, including difficulty in focusing, sweating, and blurred vision. All symptoms disappeared within a few minutes after the simulation.

Considering the level of satisfaction with the method, 100% of the respondents found the intervention interesting, 93.3% thought it was entertaining, 86.7% found it realistic and 78% considered it an important part of their rehabilitation. 77.3% sought to apply the learned skills in everyday life, and even if they didn't practice them much, 93.3% reported subjective improvement in communicational skills. During the intervention, most patients felt safe using the devices (93.3%), and 80% of them considered the communication with the avatar pleasant. On the other hand, difficulties have also been reported. Namely, 13.4% of the patients thought that the use of TDC makes the intervention cumbersome (they found it quite slow), and 20% of them complained of anxiety related to the method. Finally, 26.7% of the patients found the TDC task a difficult part of the intervention.

4.1.7. Sustainability of treatment effect

Although the generalization effect of VR-ToMIS on neurocognitive skills wasn't confirmed, the treatment-associated changes of negative ($F(1,41) = 14.67$, $p = .0005$, $\eta_p^2 = 0.29$) and cognitive symptoms ($F(1,41) = 6.94$, $p = .01$, $\eta_p^2 = 0.16$) proved to be sustainable. Moreover, at the follow-up assessment, the advantage of the experimental intervention was supported even in case of affective symptoms (i.e.: depression, anxiety, feeling of guilt) ($F(1,41) = 24.96$, $p < .0001$, $\eta_p^2 = 0.41$). Regarding the long-term effects of the training on ToM and pragmatic skills, results stayed significant, or showed a trend towards significance three months after the intervention in all cases. However, effect sizes showed a tendency to fall back into a lower range ($\eta_p^2 = 0.03–0.34$, d = 0.26–1.04). Finally, concerning quality of life, significant benefit of VR-ToMIS was supported at follow-up ($F(1,41) = 4.10$, $p = .05$, $\eta_p^2 = 0.10$) (Fig. 2.)

5. Discussion

The primary aim was to develop an intervention that can be used to improve ToM in its complexity by providing a safe learning environment where patients can experience and analyze social interactions with the help of an experienced therapist. To our knowledge, this is the first method that aims to reach this goal by using VR technology.

Due to the novelty of the use of such technology in therapies, the potential motivational effect is often mentioned in the literature, but the reported experience is mixed. [49–51] In this regard, it is worth mentioning that no patient has dropped out during the intervention, not even from the passive VR condition due to consent withdrawal. In

Table 2
Group characteristics across time.

	VR-ToMIS			Passive VR		
	T0 (n = 21)	T1 (n = 21)	T2 (n = 21)	T0 (n = 21)	T1 (n = 21)	T2 (*n* = 19)
RMET	19.14(4.15)	22.42(4.22)	22.04(3.45)	18.62(3.29)	18.90(4.66)	19.82(4.98)
PANSS	52.04(10.19)	42.76(11.62)	42.57(9.32)	51.80(9.87)	55(9.97)	52(11.3)
Positive symptoms	8.90(2.47)	10.38(14.86)	7.76(2.02)	8.90(2.47)	9.13(3.76)	8.36(2.05)
Negative symptoms	15.28(4.96)	12.28(3.92)	12.47(4.27)	15.33(3.61)	16.14(4.59)	16.15(3.86)
Cognitive symptoms	13.62(3.27)	11.66(3.02)	11.52(2.56)	14.57(3.07)	14.90(3.78)	14.15(3.82)
Activity /Excitement symptoms	4.86(1.01)	5.23(1.95)	4.57(0.97)	5.57(1.33)	5.62(1.36)	5.68(2.76)
Affective symptoms	9.19(2.63)	7.85(3.96)	6.66(2.19)	8.33(3.31)	8.76(2.19)	8.47(2.76)
RBANS total scale	88.38(17.21)	96.66(15.25)	95.66(12.96)	76.05(15.55)	77.76(14.27)	82.26(17.21)
Immediate memory	89.52(19.68)	100.47(20.28)	104.61(19.87)	71.05(19.70)	74.95(18.22)	83.52(23.11)
Attention	73.33(18.25)	81.86(14.91)	83.95(18.18)	72.57(15.43)	72.95(14.27)	78.89(11.11)
Visuospatial skills	110.43(14.83)	109.52(12.11)	108.8(12.5)	100.57(16.75)	97.24(18.10)	92.68(23.40)
Language	98.90(19.69)	99.09(14.11)	101.09(12.6)	90.95(12.56)	92.28(9.19)	93.78(12.71)
Delayed memory	91.04(17.92)	97.19(15.41)	92.09(27.68)	71.00(22.48)	76.71(14.27)	77.11(23.84)
WCST-64						
Correct answers	33.57(13.34)	37.71(13.21)	39(12.78)	33.09(13.65)	32.85(14.86)	35.57(11.84)
Perseverative errors	8.28(7.75)	8.19(7.55)	8.04(6.74)	10.38(9.26)	10.00(11.18)	11.57(7.86)
Completed categories	1.28(7.75)	1.9(1.61)	1.95(1.71)	1.76(1.58)	1.52(1.78)	1.73(1.37)
Metaphor and Irony test						
Metaphor comprehension	7.38(1.85)	8.66(1.06)	8.23(1.70)	6.00(2.85)	5.47(2.94)	5.94(2.63)
Irony comprehension	8.25(2.32)	9.62(1.16)	9.71(0.90)	6.81(3.50)	6.52(2.96)	6.73(2.91)
Quantity implicature	8.04(3.33)	11.38(4.07)	9.76(4.31)	7.81(3.38)	6.57(3.41)	6.00(2.88)
Quality implicature	7.14(3.23)	9.57(2.71)	9.76(4.31)	5.38(3.57)	3.81(3.91)	3.63(3)
Manner implicature	8.24(3.84)	13.52(2.78)	11.61(3.77)	7.85(3.32)	7.47(3.25)	7.1(2.71)
Relevance implicature	6.14(3.45)	9.76(3.56)	8.85(3.74(	6.43(2.97)	4.57(2.56)	4.94(3.16)
ToM score	18.43(7.21)	25.05(6.13)	23.14(6.19)	17.61(5.38)	15.66(6.56)	14.26(5.88)
Cartoon test						
First order ToM	3.81(1.25)	4.66(0.65)	4.52(0.81)	3.43(1.25)	3.23(1.51)	3.15(1.23)
Second order ToM	3.67(0.85)	4.61(0.67)	4.52(0.87)	3.57(1.07)	3.28(1.15)	3.42(1.06)
Third order ToM	1.14(0.96)	2.38(0.80)	2.38(1.02)	1.23(1.13)	0.95(1.07)	0.94(0.97)
Faux pas test						
Overall scores of Faux pas stories	0.53(0.24)	0.76(0.19)	0.66(0.28)	0.47(0.21)	0.42(0.24)	0.44(0.24)
Overall scores of non-Faux pas stories	0.93(0.12)	0.93(0.22)	0.96(0.07)	0.89(0.17)	0.88(0.17)	0.87(0.17)
Faux pas recognition	0.64(0.27)	0.83(0.28)	0.78(0.28)	0.60(0.28)	0.58(0.29)	0.6(0.25)
Faux pas understand	0.41(0.29)	0.69(0.29)	0.67(0.29)	0.39(0.25)	0.35(0.29)	0.4(0.27)
Inferring intention	0.32(0.25)	0.59(0.28)	0.50(0.33)	0.23(0.22)	0.26(0.24)	0.30(0.23)
Understand false belief	0.52(0.31)	0.73(0.25)	0.63(0.33)	0.35(0.25)	0.38(0.28)	0.38(0.25)
Empathy	0.42(0.29)	0.70(0.22)	0.64(0.30)	0.41(0.29)	0.29(0.28)	0.31(0.27)
LQoLP						
Work satisfaction	16.71(5.53)	16.71(9.69)	16.14(6.52)	17.57(6.06)	16.52(5.68)	15.73(6.17)
Leisure activities	15.28(3.51)	15.62(3.93)	15.09(5.03)	16.09(4.21)	16.33(4.19)	16.21(3.31)
Religion	12.09(2.75)	10.42(5.41)	10.04(5.51)	12.52(4.83)	12.14(4.33)	11.52(4.15)
Finances	10.09(2.76)	9.38(3.12)	9.52(3.64)	9.28(3.05)	9.00(3.35)	9.00(3.31)
Living situation	38.90(6.76)	38.09(8.56)	38.9(9.74)	37.00(10.16)	39.28(7.38)	37.21(8.36)
Safety	12.14(1.42)	12.00(2.02)	11.76(2.34)	10.90(4.50)	11.24(2.79)	10.78(2.37)
Family relations	16.33(4.63)	17.38(6.93)	16.42(5.41)	18.76(5.02)	19.28(5.02)	18.05(6.07)
Social relations	8.66(3.15)	9.43(2.25)	9(2.86)	10.09(3.03)	9.43(2.50)	8.26(3.41)
Health	14.95(2.89)	15.33(3.62)	16.04(2.45)	14.47(4.09)	14.71(3.13)	14.47(3.48)
Self-esteem	5.48(2.80)	6.38(2.85)	6.42(2.58)	6.00(2.51)	5.76(1.97)	5.31(2.51)
Cantril's ladder	7.08(2.63)	7.76(2.64)	7.78(2.69)	6.51(3.28)	6.76(2.85)	6.21(2.59)

summary, as the subjective feedback of the patients show, the intervention was well tolerated and received a positive evaluation. Additionally, no severe side effects were reported during the intervention. When mild symptoms were present, no treatment was required.

Regarding the main purpose of the intervention, the presented results provide preliminary evidence for the feasibility of a novel VR-based ToM intervention for schizophrenic outpatients.

While the possibility of improvement of ToM abilities in schizophrenic patients has already been confirmed, the results are usually controversial even in the case of ToM-focused methods, where the efficacy on higher-order ToM skills appears to be limited. From the perspective of development, the variability of results is often explained by the applied techniques or the short duration of the interventions [3,15]. As additional explanatory factors, sampling procedure and assessment difficulties could also be considered. To address these difficulties, a number of approaches were taken. For instance, VR technology was used to make the simulations more realistic, techniques including both affective and cognitive aspects of ToM with different levels of complexity were applied, only outpatients with detectable ToM deficits were included, and finally, a set of ToM tests covering different aspects of ToM were used. As our results show, all aspects of ToM have been improved, even higher-order ones, whereas the improvement was sustainable even 3-month after the study, supporting our hypothesis that the use of the combination of immersive-interactive technology, and cognitive therapeutic techniques might be advantageous to improve ToM skills in schizophrenic patients. However, the experienced decline in the long-term effect of the intervention warns that changes like the intensification of the training might be necessary for further sustainability.

Besides the feasibility of the method, the possible moderator role of IQ was also investigated, as literature data on the association of IQ with ToM is controversial. At first glance, even meta-analyses couldn't resolve the contradictions. However, after taking a closer look at their results, they can be separated along with the severity of the patients' condition, implying the independence of the two areas in the acute phase of schizophrenia, and their significant association when in remission [52,53]. On the other hand, to our knowledge, the effect of IQ on intervention-related ToM improvement has been examined by only

Table 3
Results of ANCOVA.

	ANCOVA (T0-T1)					ANCOVA (T0-T2)				
	F (1,41)	p	η2	d VR-ToMIS	d passive-VR	F (1,41)	p	η2	d (VR-ToMIS)	d (passive VR)
RMET	**4.93**	**0.03**	**0.12**	**0.78**	0.08	1.66	0.20	0.04	**0.64**	0.39
PANSS	**25.54**	**<0.0001**	**0.42**	**−0.71**	0.58	**22.89**	**<0.0001**	**0.39**	**−1.99**	0.02
Positive symptoms	1.05	0.31	0.02	0.10	−0.09	2.20	0.14	0.05	−0.17	−0.29
Negative symptoms	**8.67**	**0.005**	**0.19**	**−0.98**	0.17	**14.67**	**0.0005**	**0.29**	**−0.98**	0.27
Cognitive symptoms	**9.67**	**0.003**	**0.21**	**−0.98**	0.12	**6.94**	**0.01**	**0.16**	**−0.85**	−0.18
Activity /Excitement symptoms	0.86	0.36	0.02	0.19	0.04	1.45	0.23	0.03	−0.29	0.07
Affective symptoms	2.38	0.13	0.06	−0.42	0.13	**24.96**	**<0.0001**	**0.41**	**−1.28**	0.06
RBANS total scale	**7.13**	**0.01**	**0.16**	**0.74**	0.23	1.04	0.31	0.02	0.48	0.73
Immediate memory	**5.99**	**0.01**	**0.14**	**0.86**	0.36	1.40	0.24	0.03	1.08	0.94
Attention	1.16	0.28	0.03	0.59	0.02	0.10	0.75	0.002	0.72	0.40
Visuospatial skills	1.83	0.18	0.04	−0.06	−0.20	2.01	0.16	0.05	−0.12	−0.41
Language	0.00	0.95	0.0001	0.01	0.14	0.00	0.95	0.0001	0.28	0.32
Delayed memory	2.89	0.09	0.07	0.44	0.46	0.05	0.82	0.001	0.03	0.44
WCST-64										
Correct answers	0.38	0.54	0.01	0.34	−0.01	0.37	0.54	0.01	0.43	0.13
Perseverative errors	0.11	0.73	0.003	−0.01	−0.03	2.55	0.11	0.06	−0.02	0.13
Completed categories	0.04	0.84	0.001	0.34	−0.16	0.02	0.87	0.0007	0.37	−0.01
Metaphor and Irony test										
Metaphor comprehension	**20.32**	**<0.0001**	**0.36**	**0.81**	−0.19	**4.59**	**0.03**	**0.11**	**0.41**	−0.02
Irony comprehension	**16.88**	**0.0002**	**0.32**	**0.57**	−0.15	**11.92**	**0.001**	**0.25**	**0.68**	−0.02
Quantity implicature	**15.46**	**0.0004**	**0.30**	**0.90**	−0.42	**5.08**	**0.03**	**0.12**	**0.26**	−0.70
Quality implicature	**24.22**	**<0.0001**	**0.40**	**0.58**	**−0.54**	**15.76**	**0.0003**	**0.31**	**0.58**	**−0.73**
Manner implicature	**44.80**	**<0.0001**	**0.56**	**1.33**	−0.16	**14.48**	**0.0005**	**0.29**	**0.64**	−0.24
Relevance implicature	**24.17**	**<0.0001**	**0.40**	**0.81**	**−0.73**	**8.33**	**0.006**	**0.19**	**0.56**	**−0.50**
ToM score	**30.81**	**<0.0001**	**0.46**	**1.17**	−0.37	**22.35**	**<0.0001**	**0.38**	**0.66**	−0.62
Cartoon test										
First order ToM	**12.34**	**0.001**	**0.26**	**0.74**	−0.17	**14.68**	**0.0005**	**0.29**	**0.68**	−0.25
Second order ToM	**18.20**	**0.0001**	**0.34**	**0.98**	−0.23	**9.31**	**0.004**	**0.21**	**0.93**	−0.11
Third order ToM	**19.09**	**0.0001**	**0.35**	**1.19**	−0.30	**18.03**	**0.0002**	**0.34**	**0.78**	−0.22
Faux pas test										
Overall scores of Faux pas stories	**34.47**	**<0.0001**	**0.49**	**1.16**	−0.38	**6.45**	**0.01**	**0.15**	**0.72**	−0.16
Overall scores of non-Faux pas stories	0.07	0.79	0.001	−0.07	−0.07	**4.42**	**0.04**	**0.10**	**0.20**	−0.05
Faux pas recognition	**9.19**	**0.004**	**0.20**	**0.92**	−0.07	**5.83**	**0.02**	**0.14**	**0.82**	−0.006
Faux pas understand	**14.67**	**0.0005**	**0.29**	**1.15**	−0.18	**8.61**	**0.005**	**0.19**	**1.04**	0.04
Inferring intention	**11.30**	**0.001**	**0.24**	**1.15**	0.14	1.12	0.29	0.03	0.66	0.38
Understand false belief	**11.76**	**0.001**	**0.25**	**1.11**	0.16	2.32	0.13	0.06	0.38	0.17
Empathy	**20.13**	**<0.0001**	**0.36**	**0.93**	−0.38	**14.68**	**0.0005**	**0.29**	**0.84**	−0.42
LQoLP										
Work satisfaction	0.02	0.89	0.0005	0.00	−0.49	0.46	0.50	0.01	−0.17	−0.79
Leisure activities	0.18	0.67	0.005	0.11	0.09	0.03	0.86	0.0009	−0.03	0.02
Religion	4.82	0.03	0.12	−0.41	−0.13	0.92	0.34	0.02	−0.58	−0.32
Finances	0.55	0.46	0.01	−0.40	−0.10	1.11	0.30	0.03	−0.19	−0.12
Living situation	2.81	0.10	0.07	−0.15	0.45	0.63	0.43	0.01	0.00	0.02
Safety	0.37	0.54	0.01	−0.07	0.07	0.06	0.81	0.001	−0.20	−0.03
Family relations	0.00	0.97	0.0000	0.19	0.17	0.00	0.95	0.0001	0.02	−0.18
Social relations	0.99	0.32	0.02	0.33	−0.26	**4.10**	**0.05**	**0.10**	**0.20**	−0.57
Health	0.03	0.86	0.0009	0.10	0.09	1.60	0.21	0.04	0.56	0.00
Self-esteem	1.40	0.24	0.03	0.34	−0.10	**3.78**	**0.05**	**0.09**	**0.36**	−0.32
Cantril's ladder	0.06	0.80	0.001	0.29	0.09	1.73	0.19	0.04	0.26	−0.14

Table 4
Impact of the intervention according to family, friends, and colleagues. ($n = 12$.

		Much worse	Worse	Not changed	Better	Much better
Conversational skills	Initiative in conversation with relatives	0%	0%	16.70%	58.30%	25.0%
	Willingness to get involved in conversations with relatives and friends	0%	0%	16.70%	66.70%	16.70%
	Willingness to communicate with unknown people	0%	0%	25%	56,25%	18,75%
	Following conversational rules	0%	0%	25%	50%	25%
Misunderstanding mental state	Misunderstanding others' intentions	0%	8.30%	25%	42%	25%
	Degree of general mistrust	0%	8.30%	50%	33.30%	8.30%
Degree of independence	Extent of activity around the house	0%	0%	33.30%	50%	16.70%
	Willingness to leave the house	0%	0%	41.70%	33.30%	25%
	Ability to travel independently	0%	0%	58.30%	0%	41.70%
	Willingness to act independently	0%	0%	50%	50%	0%
	Extent of autonomy	0%	0%	58.30%	17%	25%

two studies so far [15,54]. Only one of them concluded that lower IQ might be limiting the beneficial effect of an intervention [53]. According to our results, in the case of lower-order ToM, the improvement was independent of IQ. However, the effect on higher-order ToM skills was moderate, suggesting a compensatory mechanism linked to cognitive skills. This also partially confirms the results by Marsh et al. (2013),

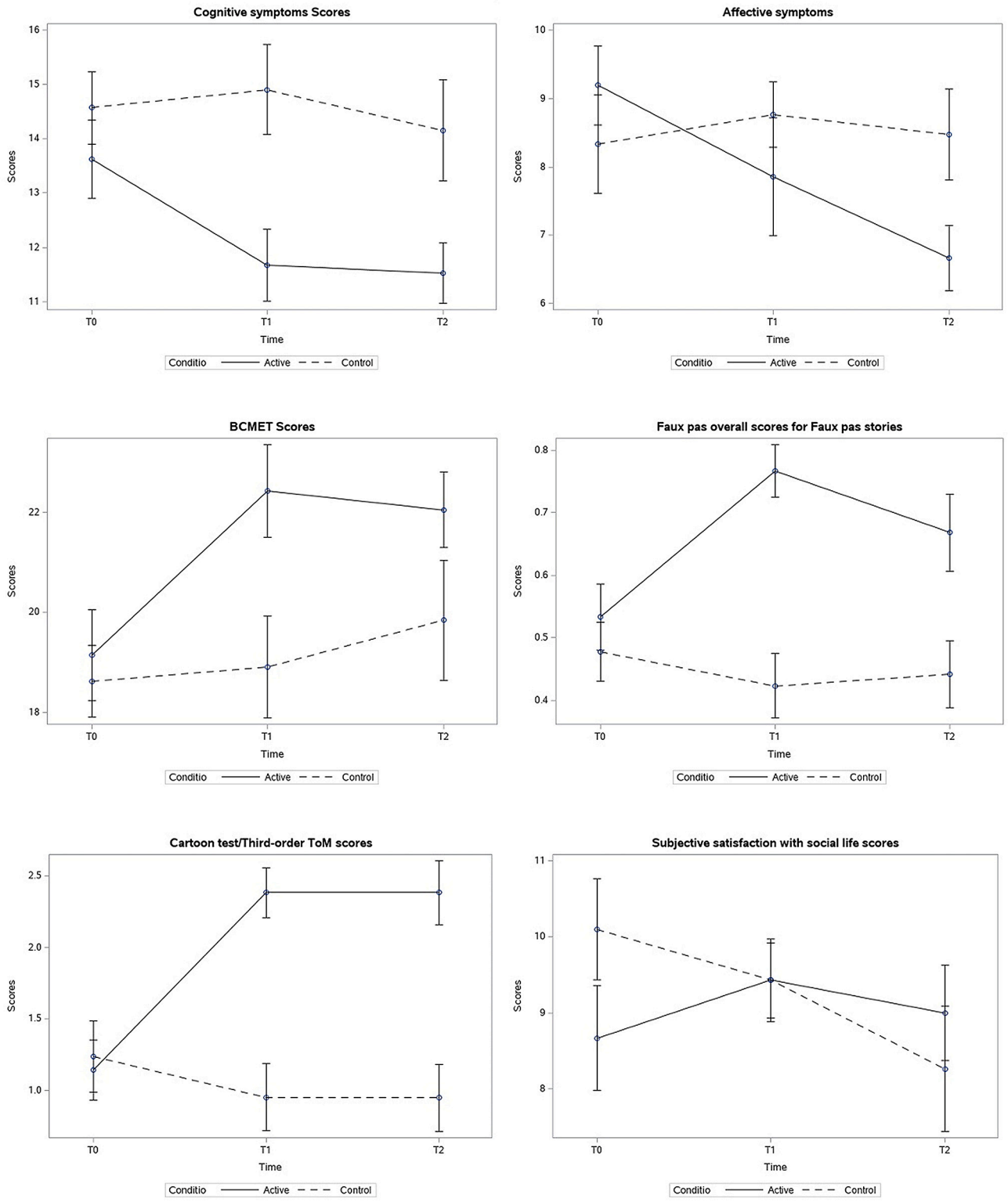

Fig. 2. Efficacy of VR-ToMIS on scores over time.

implying that lower IQ might reduce the beneficial effect of VR-ToMIS on higher-order mental state attribution [54].

The relationship between pragmatics and ToM has also been frequently discussed and has been a subject of debate for years [37,55]. Clarification of the issue is important, not only from a theoretical aspect but also from a practical point of view, since Bosco (2018) highlighted that researchers often use assessment tools that can measure both constructs, risking the validity of their conclusion. [37] To avoid methodological issues, we chose to assess pragmatics separately. Based on our results, modifiability was supported in the case of all covered implicatures, accompanied by the long-term sustainability of the improvement. Although the results indicate that - at least in the case of the detection of relevance and manner implicature - the improvement might be affected by IQ, considering the literature, a separate

examination of verbal and non-verbal aspects of IQ may lead to a more accurate conclusion on the relationship in future studies. [56]

Generalization effect on symptoms and neurocognitive skills has also been analyzed. Many studies have examined the possible generalization of social cognitive treatment effects on symptomatology. However, in most cases (even when improvement in overall symptoms was reported) advantageous effects on positive or negative symptoms weren't supported [20]. In contrast, in case of our study, significant long-term improvement was found in the case of negative, cognitive, and affective symptoms of schizophrenia. However, it must be mentioned that instead of using the usual method, PANSS scoring was based on a recent meta-analytic exploration of the scale, suggesting redefining the symptom groups [25]. Regarding neurocognition, even though it wasn't a treatment target of VR-ToMIS, the between-group comparison showed significant improvement after the treatment in the case of immediate memory. However, this effect wasn't detectable three-month after the intervention. Unfortunately, only a few studies have investigated the possible effects of social cognitive training on neurocognition, but in line with our experiences, their results imply that such methods may have an advantageous effect on memory as a direct result of these interventions, yet the sustainability of the results wasn't studied. [19,57]

As for the effect on functionality, the meta-analysis of Kurtz and Richardson (2012) supported the generalization effect of social cognitive interventions on social functioning. The association was confirmed by using observer-rated assessments. On the other hand, recent analyses, based on self-administered and performance-based QoL assessments led to contradictory results. Inconsistency in the reported results can be explained by many factors, such as the diversity of assessment tools [20,57,59]. Most measurements are based on mainly permanent fields, the factors of which can hardly be changed by a few weeks of training. With this in mind, the lack of significant change on LQoLP scales after the intervention is not surprising. However, in line with the analysis by Kurtz and Richardson (2012), the acquaintances of the patients interviewed after the training reported noticeable changes in the level of social functioning, particularly in the case of communication skills. [20] Furthermore, significant advantageous changes in the subjective evaluation of social relationships at follow-up suggest that patients might successfully transfer the learned skills to real-life interactions. Despite the encouraging experiences, in order to assess the stability of the effects, a better-powered longitudinal study is needed.

Finally, some limitations should be considered when interpreting the results. First, although our efforts to use multiple assessment tools to cover a wide range of possibly relevant factors can be considered a strength of our study, it was also an important limitation, as multiple testing may raise the Type I error rate, as well as its tiring nature, which has already resulted in a 20% ($n = 11$) dropout rate on its own. However, due to the lack of sufficiently comprehensive ToM measurements, our possibility to reduce the number of assessments was limited. Second, as a result of the strict inclusion and exclusion criteria, the screen failure rate reached 57.84%, leading to a small sample size, therefore limiting the generalizability of the results and the power of the study. The problem of generalizability is somewhat reduced by the fact that the currently presented study confirms the results of a recently published case report and a pilot study on the feasibility of VR-ToMIS. Nevertheless, to understand the effects of the interventions it is important to confirm the presented results in a better-powered long-term study on a larger sample size and with an active control group.

Despite the mentioned limitations and the preliminary nature of our results, they support the potential of the integration of modern technology and traditional methods for future interventions.

Funding

The study was supported by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Neurology thematic program of Semmelweis University.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declaration of Competing Interest

None.

Acknowledgment

The authors thank the staff of the Moravcsik Foundation, Budapest Art Brut Gallery, and Ébredés Foundation for their assistance with recruitment.

References

[1] Charrier N, Chevreul K, Durand-Zaleski I. The cost of schizophrenia: a literature review. Enceph Psychiatr Clin Biol Ther 2013;39:S49–56. https://doi.org/10.1016/j.encep.2012.11.004.
[2] Bechi M, Spangaro M, Pigoni A, Ripamonti E, Buonocore M, Cocchi F, et al. Exploring predictors of work competence in schizophrenia: the role of theory of mind. Neuropsychol Rehabil 2017;29:619–703. https://doi.org/10.1080/09602011.2017.1314217.
[3] Vass E, Fekete Z, Simon V, Simon L. Interventions for the treatment of theory of mind deficits in schizophrenia: systematic literature review. Psychiatry Res 2018;267:37–47. https://doi.org/10.1016/j.psychres.2018.05.001.
[4] Brüne M. Theory of mind and social competence in schizophrenia. Clin Neuropsychiatry 2006;3:132–8.
[5] Pinkham AE, Penn DL. Neurocognitive and social cognitive predictors of interpersonal skill in schizophrenia. Psychiatry Res 2006;143:167–78. https://doi.org/10.1016/j.psychres.2005.09.005.
[6] Frith CD. Schizophrenia and theory of mind. Psychol Med 2004;34:385–9. https://doi.org/10.1017/S0033291703001326.
[7] Abu-Akel A, Shamay-Tsoory SG. Characteristics of theory of mind impairments in schizophrenia. Soc Cogn Schizophr From Evid Treat 2013:196–214. https://doi.org/10.1093/med:psych/9780199777587.003.0008.
[8] Brüne M, Brüne-Cohrs U. Theory of mind-evolution, ontogeny, brain mechanisms and psychopathology. Neurosci Biobehav Rev 2006;30:437–55. https://doi.org/10.1016/j.neubiorev.2005.08.001.
[9] Lysaker PH, Leonhardt BL, Brüne M, Buck KD, James A, Vohs J, et al. Capacities for theory of mind, metacognition, and neurocognitive function are independently related to emotional recognition in schizophrenia. Psychiatry Res 2014;219:79–85. https://doi.org/10.1016/j.psychres.2014.05.004.
[10] Kalbe E, Schlegel M, Sack AT, Nowak DA, Dafotakis M, Bangard C, et al. Dissociating cognitive from affective theory of mind: a TMS study. Cortex 2010;46:769–80. https://doi.org/10.1016/j.cortex.2009.07.010.
[11] Sebastian CL, Fontaine NMG, Bird G, Blakemore S-J, De Brito SA, McCrory EJP, et al. Neural processing associated with cognitive and affective theory of mind in adolescents and adults. Soc Cogn Affect Neurosci 2012;7:53–63. https://doi.org/10.1093/scan/nsr023.
[12] Bosia M, Riccaboni R, Poletti S. Neurofunctional correlates of theory of mind deficits in schizophrenia. Curr Top Med Chem 2012;12:2284–302. https://doi.org/10.2174/156802612805289917.
[13] Mazza M, De Risio A, Surian L, Roncone R, Casacchia M. Selective impairments of theory of mind in people with schizophrenia. Schizophr Res 2001;47:299–308. https://doi.org/10.1016/S0920-9964(00)00157-2.
[14] Galderisi S, Rucci P, Kirkpatrick B, Mucci A, Gibertoni D, Rocca P, et al. Interplay among psychopathologic variables, personal resources, context-related factors, and real-life functioning in individuals with schizophrenia a network analysis. JAMA Psychiat 2018;75:396–404. https://doi.org/10.1001/jamapsychiatry.2017.4607.
[15] Bechi M, Spangaro M, Bosia M, Zanoletti A, Fresi F, Buonocore M, et al. Theory of mind intervention for outpatients with schizophrenia. Neuropsychol Rehabil 2013;23:383–400. https://doi.org/10.1080/09602011.2012.762751.
[16] Bechi M, Bosia M, Spangaro M, Buonocore M, Cocchi F, Pigoni A, et al. Combined social cognitive and neurocognitive rehabilitation strategies in schizophrenia: neuropsychological and psychopathological influences on theory of mind improvement. Psychol Med 2015;45:3147–57. https://doi.org/10.1017/S0033291715001129.
[17] Bechi M, Riccaboni R, Ali S, Fresi F, Buonocore M, Bosia M, et al. Theory of mind and emotion processing training for patients with schizophrenia: preliminary findings. Psychiatry Res 2012;198:371–7. https://doi.org/10.1016/j.psychres.2012.02.004.
[18] Mazza M, Lucci G, Pacitti F, Pino MC, Mariano M, Casacchia M, et al. Could schizophrenic subjects improve their social cognition abilities only with observation and imitation of social situations? Neuropsychol Rehabil 2010;20:675–703. https://doi.org/10.1080/09602011.2010.486284.

[19] Veltro F. A comparison of the effectiveness of problem solving training and of cognitive-emotional rehabilitation on neurocognition, social cognition and social functioning in people with schizophrenia. Clin Pract Epidemiol Ment Health 2011; 7:123–32. https://doi.org/10.2174/1745017901107010123.
[20] Kurtz MM, Richardson CL. Social cognitive training for schizophrenia: a meta-analytic investigation of controlled research. Schizophr Bull 2012;38:1092–104. https://doi.org/10.1093/schbul/sbr036.
[21] Pino MC, Pettinelli M, Clementi D, Gianfelice C, Mazza M. Improvement in cognitive and affective theory of mind with observation and imitation treatment in subjects with schizophrenia. Clin Neuropsychiatry 2015;12:64–72.
[22] Vass E, Fekete Z, Lencse L, Ecseri M, Kis B, Szekrényes-Varga Á, et al. Treatment of theory of mind deficits in schizophrenia by using virtual reality, the VR-ToMIS training (virtual reality based theory of mind intervention in schizophrenia). Psychiatr Hung 2019;34:287–99.
[23] Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. J Child Psychol Psychiatry 2001;42:241–51. https://doi.org/10.1111/1469-7610.00715.
[24] Barrera Á, Vázquez G, Tannenhaus L, Lolich M, Herbst L. Theory of mind and functionality in bipolar patients with symptomatic remission. Rev Psiquiatr Salud Ment (English Ed) 2013;6:67–74. https://doi.org/10.1016/j.rpsmen.2012.07.003.
[25] Shafer A, Dazzi F. Meta-analysis of the positive and negative syndrome scale (PANSS) factor structure. J Psychiatr Res 2019;115:113–20. https://doi.org/10.1016/j.jpsychires.2019.05.008.
[26] Schmank CJ, Goring SA, Kovacs K, Conway ARA. Psychometric network analysis of the Hungarian WAIS. J Intelligence 2019;7:21. https://doi.org/10.3390/jintelligence7030021.
[27] Dong Y, Thompson CL, Huey Joanne Tan S, Ben Swie Lim L, Pang W, Li-Hsian Chen C. Test-retest reliability, convergent validity and practice effects of the RBANS in a memory clinic setting: a pilot study. Open J Med Psychol 2013;02:11–6. https://doi.org/10.4236/ojmp.2013.24B003.
[28] Randolph C, Tierney MC, Mohr E, Chase TN. The repeatable battery for the assessment of neuropsychological status (RBANS): preliminary clinical validity. J Clin Exp Neuropsychol 1998;20:310–9. https://doi.org/10.1076/jcen.20.3.310.823.
[29] Erika N, Francisco B. The Wisconsin card sorting test and the cognitive assessment of prefrontal executive functions: A critical update. Brain Cogn 2009;71:437–51. https://doi.org/10.1016/j.bandc.2009.03.005.
[30] Greve KW. The WCST-64: a standardized short-form of the Wisconsin card sorting test. Clin Neuropsychol 2001;15:228–34. https://doi.org/10.1076/clin.15.2.228.1901.
[31] Bowden SC, Fowler KS, Bell RC, Whelan G, Clifford CC, Ritter AJ, et al. The reliability and internal validity of the Wisconsin card sorting test. Neuropsychol Rehabil 1998;8:243–54. https://doi.org/10.1080/713755573.
[32] Fernández-Abascal EG, Cabello R, Fernández-Berrocal P, Baron-Cohen S. Test-retest reliability of the "reading the mind in the eyes" test: a one-year follow-up study. Mol Autism 2013;4:33. https://doi.org/10.1186/2040-2392-4-33.
[33] Fernández-Modamio M, Arrieta-Rodríguez M, Bengochea-Seco R, Santacoloma-Cabero I, Gómez de Tojeiro-Roce J, García-Polavieja B, et al. Faux-pas test: a proposal of a standardized short version. Clin Schizophr Relat Psychoses 2018. https://doi.org/10.3371/CSRP.FEAR.061518. CSRP.FEAR.061518.
[34] Varga E, Tényi T, Fekete S, Herold R. The evaluation of mentalization deficit by the faux pas test in schizophrenia. Neuropsychopharmacol Hung 2008;10:75–80.
[35] Fekete Z, Vass E, Balajth R, Tana Ü, Nagy AC, Domján N, et al. Regrouping scalets: psychometric properties of the theory of mind picture stories task in a schizophrenic sample. Neuropsychol Rehabil 2021:1–21. https://doi.org/10.1080/09602011.2021.1930559.
[36] Varga E, Simon M, Tényi T, Schnell Z, Hajnal A, Orsi G, et al. Irony comprehension and context processing in schizophrenia during remission - a functional MRI study. Brain Lang 2013;126:231–42. https://doi.org/10.1016/j.bandl.2013.05.017.
[37] Bosco FM, Tirassa M, Gabbatore I. Why pragmatics and theory of mind do not (completely) overlap. Front Psychol 2018;13:1459. https://doi.org/10.3389/fpsyg.2018.01453.
[38] Gaite L, Vázquez-Barquero JL, Arriaga Arrizabalaga A, Schene AH, Welcher B, Thornicroft G, et al. Quality of life in schizophrenia: development, reliability and internal consistency of the Lancashire quality of life profile - European version. EPSILON study 8. Br J Psychiatry 2000;29:s49–54. https://doi.org/10.1192/bjp.177.39.s49.
[39] van Nieuwenhuizen C, Schene AH, Koeter MWJ, Huxley PJ. The Lancashire quality of life profile: modification and psychometric evaluation. Soc Psychiatry Psychiatr Epidemiol 2001;36:36–44. https://doi.org/10.1007/s001270050288.
[40] Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993;3:2013–220. https://doi.org/10.1207/s15327108ijap0303_3.
[41] Sevinc V, Berkman MI. Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. Appl Ergon 2020;82:102958. https://doi.org/10.1016/j.apergo.2019.102958.
[42] Csukly G, Simon L, Kiss B, Takács B. Evaluating psychiatric patients using high fidelity animated faces. Annu Rev Cyber Ther Telemed 2004;7:278–9.
[43] Vass E, Simon V, Fekete Z, Lencse L, Ecseri M, Kis B, et al. A novel virtual reality-based theory of mind intervention for outpatients with schizophrenia: A proof-of-concept pilot study. Clin Psychol Psychother 2020. https://doi.org/10.1002/cpp.2519. cpp.2519.
[44] Shieh G. Power analysis and sample size planning in ANCOVA designs. Psychometrika 2020;85:101–20. https://doi.org/10.1007/s11336-019-09692-3.
[45] Khammar A, Yarahmadi M, Madadizadeh F. What is analysis of covariance (ANCOVA) and how to correctly report its results in medical research? Iran J Public Health 2020;49:1016–7.
[46] Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014;34:502–8. https://doi.org/10.1111/opo.12131.
[47] Aaker DA, Keppel G. Design and analysis: a researcher's handbook. J Market Res 1976;12:318. https://doi.org/10.2307/3150757.
[49] Dellazizzo L, Potvin S, Luigi M, Dumais A. Evidence on virtual reality–based therapies for psychiatric disorders: meta-review of meta-analyses. J Med Internet Res 2020;22:e20889. https://doi.org/10.2196/20889.
[50] Park KM, Ku J, Choi SH, Jang HJ, Park JY, Kim SI, et al. A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. Psychiatry Res 2011;189:166–72. https://doi.org/10.1016/j.psychres.2011.04.003.
[51] Pot-Kolder R, Geraets C, Veling W, van Beilen M, Staring A, Gijsman H, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018;5:217–26. https://doi.org/10.1016/S2215-0366(18)30053-1.
[52] Bora E, Yucel M, Pantelis C. Theory of mind impairment in schizophrenia: Meta-analysis. Schizophr Res 2009;109:1–9. https://doi.org/10.1016/j.schres.2008.12.020.
[53] Sprong M, Schothorst P, Vos E, Hox J, Van Engeland H. Theory of mind in schizophrenia: Meta-analysis. Br J Psychiatry 2007. https://doi.org/10.1192/bjp.bp.107.035899.
[54] Marsh P, Langdon R, McGuire J, Harris A, Polito V, Coltheart M. An open clinical trial assessing a novel training program for social cognitive impairment in schizophrenia. Australas Psychiatry 2013;21:122–6. https://doi.org/10.1177/1039856213475683.
[55] Sperber D, Wilson D. Pragmatics, modularity and mind-reading. Mind Lang 2002; 17:3–23. https://doi.org/10.1111/1468-0017.00186.
[56] Volkers N. Diverging views on language disorders. ASHA Lead 2018;23:44–53. https://doi.org/10.1044/leader.FTR1.23122018.44.
[57] Gaudelus B, Virgile J, Geliot S, Franck N, Dupuis M, Hochard C, et al. Improving facial emotion recognition in schizophrenia: a controlled study comparing specific and attentional focused cognitive remediation. Front Psych 2016;7:105. https://doi.org/10.3389/fpsyt.2016.00105.
[59] Yeo H, Yoon S, Lee J, Kurtz MM, Choi K. A meta-analysis of the effects of social-cognitive training in schizophrenia: the role of treatment characteristics and study quality. Br J Clin Psychol 2022;61:37–57. https://doi.org/10.1111/bjc.12320.