Computers in Human Behavior 79 (2018) 40–52

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Training in a comprehensive everyday-like virtual reality environment compared to computerized cognitive training for patients with depression

Lorenz B. Dehn [a, *], Leona Kater [b], Martina Piefke [c], Mario Botsch [d], Martin Driessen [d], T. Beblo [a]

[a] Clinic for Psychiatry and Psychotherapy Bethel, Remterweg 69-71, D-33617 Bielefeld, Germany
[b] Clinic of Internal and Geriatric Medicine, Schildische Str. 99, D-33611 Bielefeld, Germany
[c] Department of Psychology and Psychotherapy, Witten/Herdecke University, Alfred-Herrhausen-Str. 50, D-58448 Witten, Germany
[d] Computer Graphics and Geometry Processing, Bielefeld University, P.O. Box 10 01 31, D-33501 Bielefeld, Germany

ARTICLE INFO

Article history:
Received 17 August 2016
Received in revised form
19 September 2017
Accepted 15 October 2017
Available online 17 October 2017

Keywords:
Virtual reality
Computerized training
Cognitive impairment
Cognitive remediation
Depression
Ecological validity

ABSTRACT

Neurocognitive impairments in patients with depression compromise everyday functioning. Thus, should neuropsychological therapy be designed as real-life-like as possible to maximize transfer effects? We investigated whether ecological validity of computerized cognitive training could be increased by a comprehensive everyday-life-simulating training device combining virtual reality, 360°-all-around visibility and autonomous navigation motions. In an eight days training program, patients exercised the learning and purchasing of shopping list products in a virtual supermarket using either the novel training device (n = 21) or a corresponding desktop application (n = 17). In a pre-post-design, effects of the two training conditions were compared regarding several outcome measures. Altogether, results did not prove a benefit of the more naturalistic training setting regarding different training performances (recognition, performance speed, spatial orientation), self-perceived daily cognitive impairments, a real-life shopping task as well as various neuropsychological capabilities. Findings are discussed in the context of general challenges in striving after ecological validity in neuropsychology.

1. Introduction

1.1. Neurocognitive impairments in depression

Besides emotional, physical or motivational dysfunctions, neurocognitive impairment is a core feature of depression (Bortolato, Carvalho, & McIntyre, 2015; Rock, Roiser, Riedel, & Blackwell, 2014). In the DSM-V (American Psychiatric Association, 2013) the impaired capacity to think, to concentrate or to make decisions is one diagnostic criterion for major depressive episode (MDD). Reviews and meta-analysis have underlined neurocognitive impairments in patients suffering from depression, especially in domains of memory, executive functions, attention and psychomotor speed (Beblo, Sinnamon, & Baune, 2011; Bora, Harrison, Yücel, & Pantelis, 2013; Lee, Hermens, Porter, & Redoblado-Hodge, 2012). In a prospective study, Conradi, Ormel, and De Jonge (2011) found that up to 66% of the patients suffered from neurocognitive impairments during a three year course of major depressive disorder. Moreover, these neurocognitive impairments remained present in up to 44% of the finally remitted patients.

It is well known that cognitive deficits in depressive patients could have important psychosocial as well as clinical implications. For instance, they significantly impair patients' social and occupational life (Evans, Iverson, Yatham, & Lam, 2014; McIntyre et al., 2013) and are related to disturbed everyday functioning (Harvey, 2011; Lee et al., 2015). Neurocognitive deficits in depression often attenuate patients adherence (e.g. through memory difficulties) and, therefore compromise effective treatment of the disorder (Papakostas, 2014). In addition, neurocognitive deficits have also found to be associated with increased suicidality (Keilp et al., 2013; Richard-Devantoy et al., 2012).

* Corresponding author. Division of Research Clinic of Psychiatry and Psychotherapy Bethel, Ev. Hospital Bielefeld, Remterweg 69-71, 33617 Bielefeld, Germany.
E-mail address: lorenz.dehn@evkb.de (L.B. Dehn).

https://doi.org/10.1016/j.chb.2017.10.019
0747-5632/

1.2. Cognitive rehabilitation

Given the notable relevance of neurocognitive deficits in persons with depression, effective intervention strategies are needed (cf. Rock et al., 2014). Among non-pharmacological treatment approaches targeting neurocognitive impairments, computerized training programs are often used (see reviews by Coyle, Traynor, & Solowij, 2015; Gates, Sachdev, Singh, & Valenzuela, 2011; Leung et al., 2015). In psychiatry, the effectiveness of computer-assisted cognitive trainings have been proven most notably for patients with schizophrenia (Grynszpan et al., 2011; Kurtz, Moberg, Gur, & Gur, 2001; Wykes, Huddy, Cellard, McGurk, & Czobor, 2014), but it has also been found to be effective in further diagnoses (Bossert, Weisbrod, & Aschenbrenner, 2014; Choi & Medalia, 2005; Fals-Stewart & Lam, 2010; Lindenmayer et al., 2008; Tchanturia et al., 2008; Vocci, 2008). For example, some initial studies indicate that computerized rehabilitation programs improve the cognitive performance of depressed patients (Bowie, Gupta, & Holshausen, 2013; Porter, Bowie, Jordan, & Malhi, 2013).

1.3. Ecological validity

As the particular relevance of ecological validity in neuropsychology has been repeatedly emphasized (Chaytor & Schmitter-Edgecombe, 2003; Sbordone, 2008), generalizability and transfer effects were also claimed for computerized cognitive training programs (Jak, Seelye, & Jurick, 2013). Lopresti, Mihailidis, and Kirsch (2004, p. 31) stated: “… *the ecological validity of ATC (assistive technology for cognition) interventions will have to be established, since it is unclear whether being able to perform a task in a controlled environment will generalize to performing the identical task in a community environment.*“. Similarly, Hampstead, Gillis, and Stringer (2014) describe the ”ultimate goal of any rehabilitation program“ in improving functioning in everyday life. Existing standard cognitive training programs often provide realistic images and scenarios. However, they are presented on a two-dimensional PC screen and, therefore, lack everyday-realism (Bryck & Fisher, 2012). This may be relevant especially in patients with depressive disorders, because they frequently claim pronounced cognitive problems in their daily living (Lahr, Beblo, & Hartje, 2007; Mowla et al., 2007).

1.4. Virtual reality

One methodology that has been increasingly considered as potential aids in enhancing the ecological validity of cognitive rehabilitation is virtual reality (VR). The various technical opportunities of VR seem to be best suited to both create naturalistic settings as well as to implement standardized procedures in treatment settings (Parsey & Schmitter-Edgecombe, 2013; Rose, Brooks, & Rizzo, 2005). By designing virtual environments that not only “look like” the real world, but actually incorporate challenges that require “real-world” functional behaviors, the ecological validity of cognitive rehabilitation is assumed to be enhanced (Rizzo, Schultheis, Kerns, & Mateer, 2004).

1.4.1. Immersion and presence

The computer-generated simulations of VR-technology engender the impression to the user as if he or she interacts within the real world while concurrently operating with or inside technical appliances (Schultheis, Himelstein, & Rizzo, 2002). Thus VR can be viewed as an advanced form of computer interface that allows the user to “interact” with and become “immersed” in a computer-generated environment (Rizzo, Buckwalter, Neumann, Kesselmann & Thiebaux, 1998). Consequently, a variety of VR approaches are available providing a continuum of experiences of immersion – the level to which the technology itself provides a comprehensible, vivid virtual environment whereby the user feels involved like “being there” (Bohil, Alicea, & Biocca, 2011; Tarr & Warren, 2002). Higher levels of presence (“being there”) will be evoked by sophisticated virtual environments that produce greater senses of immersion compared to less complex VR setups (Cummings & Bailenson, 2015; Witmer & Singer, 1998). For instance, Gorini, Capideville, De Leo, Mantovani, and Riva (2011) found, that students who experienced a virtual environment presented at an external laptop screen (low degree of immersion) stated lower levels of presence compared to those who used a highly immersive *Head-Mounted Display* (HMD), a display device, worn on the head or as part of a helmet, which fully covers the user's eyes. In this latter highly immersive condition, characters, objects and the environment itself, were perceived as more real, and the experience was judged more interesting and involving than in the low immersive laptop-screen condition. In a study of Juan and Pérez (2009) the use of a HMD-system was compared to the application of an even more immersive system called *Cave Automatic Virtual Environment* (CAVE), where the user is surrounded by the virtual reality within a large stereoscopic room. The results showed, that this fully-immersive setup induced a higher degree of presence in users. In such highly immersive VR-environments, like CAVE or HMD, “… *users are no longer simply external observers of images on a computer screen but are active participants within a computer-generated three-dimensional virtual world*” (Riva, Mantovani, & Gaggioli, 2004, p. 2). Examining the aspect of active participation within VR-applications, Freeman, Lessiter, Pugh, and Keogh (2005) gave their participants either the opportunity to autonomously navigate through a virtual environment or told them to passively keep their eyes open while the experimenter navigated the route in accordance to the other group. Subsequently, participants who self-navigated gave significantly higher engagement ratings on a presence questionnaire than persons exposed to the same VR-experience but who did not self-navigate. Diemer, Alpers, Peperkorn, Shiban, and Mühlberger (2015) hypothesized, that the subjective experience of presence in virtual reality is based on the VR-setup's level of immersion as well as the degree of arousal the user feels during his individual VR experience.

1.4.2. Multisensory learning

Depending on its features and complexity, a VR environment artificially creates sensory experiences and induces multisensory feedback, e.g. auditory, visual and proprioceptive. This could bee viewed as an underlying mechanism of VR in facilitating clinical treatments, because multisensory-training scenarios were attributed to better approximate ecological settings and to be more effective for learning, since it is *“likely that the human brain has evolved to develop, learn and operate optimally in multisensory environments”* (Shams & Seitz, 2008, p. 1).

Moreover, engaging users in a multisensory VR training environment does not only encourage active participation and involvement, but rather improves neuroplastic changes of the brain (Teo et al., 2016). Precisely, learning and practicing skills in the sense of physical activity or mental stimulation, which are key features of immersive VR applications, have found to be critical for inducing a *training-dependent reconfiguration of brain networks* (Foster, 2015) and *experience-dependent neuroplasticity*, respectively (Kleim & Jones, 2008).

Another idea that has been hypothesized as a potential benefit of VR in treatment applications concerns the activation of implicit procedural memory (Rizzo, Buckwalter, Neumann, Kesselman, & Thiebaux, 1998, 2004). This idea originally rest upon the observation of persons with neurologically based memory impairment, in which, however, procedural or skill memory capabilities often remain relatively unimpaired. This involves the capacity to learn rule-based or automatic procedures and can be contrasted to declarative memory. Here, interactive and immersive virtual reality applications are thought to provide training environments that encourage cognitive and functional improvement by tying on a person's preserved procedural abilities. Hence, Rizzo et al. (2004) presumed that, "*cognitive processes could be restored via procedures practised repetively within a virtual environment that contains functional real-world demands*" (p. 218).

1.4.3. Applications of VR

Virtual reality has increasingly proven to be an effective tool in neuroscience research as well as treatment settings for neurological and psychiatric patients (see reviews by Bohil et al., 2011; Rose et al., 2005; Parsons, 2015). Regarding neurorehabilitation, there are a lot of studies of the therapeutic use of VR in specific impairments resulting from brain injury, particularly recovery of function after stroke (Laver, George, Thomas, Deutsch, & Crotty, 2015), but also balance disorders, spatial ability impairments, visual neglect or certain cognitive dysfunctions (Rose et al., 2005). Neurorehabilitation using VR interventions offers the opportunity to expose patients to controlled settings under a range of stimulus conditions that are not easily controllable in the real world (Rizzo et al., 2004).

In neuropsychology VR has already been shown to enable an ecologically valid assessment of cognitive functions in neurological as well as psychiatric patients (Parsons, 2011; Rizzo et al., 2004; Rose et al., 2005). Many different virtual environments have yet been created referring to relevant everyday situations, for example supermarkets, grocery store, kitchens, office, apartment, park, library or street crossing; and applications that use driving simulators (Knight & Titov, 2009; Parsons, 2015; Rizzo et al., 2004). However, neuropsychological rehabilitation applications using VR still remain under development, especially in psychiatric settings. Whereas the majority of VR-based interventions in neurorehabilitation were used to train motor deficits or activities of daily living, the application as neuropsychological training scenarios for cognitive and meta-cognitive deficits has also been shown in several studies (Weiss, Kizony, Feintuch, & Katz, 2006). In their recent systematic review on the use of VR for cognitive rehabilitation after brain injury, Shin and Kim (2015) found VR programs to lead to an improvement in cognitive function, especially in the areas of memory and attention but not executive functions.

Regarding psychiatric settings, some few studies showed cognitive rehabilitation programs using VR-based setups to offer the potential for significant improvements in cognitive function compared to control conditions, e.g. in patients with schizophrenia (Marques, Queiros & Rocha, 2008), older adults with chronic schizophrenia (Chan et al., 2010), memory impaired elderly adults (Optale et al., 2010) or children with problems of inattention and impulsiveness (Cho et al., 2004).

1.4.4. Virtual reality application in patients with depression

To our knowledge, only four studies address the application of VR in patients with depressive disorders. Falconer et al. (2014) used an immersive VR scenario, in which depressed patients practiced delivering and receiving compassion using virtual embodiment. Reductions in depressions severity and self-criticism as well as increased self-compassion were gained after three intervention sessions. Two further studies found depressed patients performed worse than controls on VR navigation measures of visual-spatial memory (Cornwell et al., 2010; Gould et al., 2007), whereas another one revealed no differences in terms of neuropsychological performance during a VR spatial navigation task (Hviid et al., 2010).

1.5. Aim of this study

Some initial studies showed that computerized rehabilitation approaches can improve cognitive performance of MDD patients (see above). However, the transfer of computerized cognitive training programs to real-world environments in general has been critically discussed (see Bryck & Fisher, 2012; Jak et al., 2013). Thus, we were interested in optimizing the presentation of a computerized training program for MDD patients by the use of VR and its real-world-like features. More precisely, we wanted to investigate, whether MDD patients profit from a comprehensive reality-oriented VR-environment (highly immersive) to a greater extent than from a typical PC desktop VR-application (less immersive). In fact, it has repeatedly been suggested to assess the incremental benefit of VR over already existing methods (Rizzo et al., 2004; Teo et al., 2016). We hypothesize that our highly immersive training setup leads to an improvement of transfer effects of cognitive training on a) everyday-related cognitive and b) functional outcomes (cognitive complaints in everyday-life, real-life performance) as well as on c) neuropsychological measures.

2. Methods

2.1. Participants

Patients with depressive disorders were recruited at the Clinic of Psychiatry and Psychotherapy Bethel (Bielefeld, Germany) during their inpatient-treatment. Inclusion criteria for study participation were as follows: diagnosis of MDD according to DSM-IV, age >18 years, at least 14 days of a hospital stay (training duration) and written informed consent prior to participation. Exclusion took place in case of a MDD with psychotic symptoms, neurological disorder with central nervous system involvement or severe medical illness. All patients who volunteered to participate were consecutively assigned to the training conditions. Of the 45 initially enrolled patients, seven had to be excluded after basic diagnostics due to coexisting neurological disorders ($n = 3$) or finally non-confirmed diagnoses ($n = 4$). At last, 21 patients participated in cognitive training using the novel VR-environment, whereas 17 patients were exercising in the desktop-condition. Participants of the two groups were comparable according to sociodemographic, clinical and VR-related variables (all $p > 0.05$; see Table 1).

The study protocol was approved by the ethics committee of the University of Muenster, Germany.

2.2. Virtual-reality training device

As part of the interdisciplinary research project CITmed (Cognitive Interaction Technology in Medicine), the novel VR-device –*OctaVis* - was developed for diagnostic and rehabilitation purposes in psychiatry and neurology (Dyck et al., 2012; Grewe et al., 2013). In the OctaVis, the user is fully surrounded by eight 26″ LCD-touch-screens, on which by means of VR a typical everyday situation - grocery shopping - is simulated. Navigation through the supermarket is based on autonomous body motions using a swivel chair (to align the 360°-view walking direction) and a joystick (to adjust locomotion and speed). Selecting articles is done by real-world-like movements, as users had to reach out their arms and tap target objects on the touch screens. To ensure familiarization with this novel training-device, participants obtained detailed introductions and initial tryouts. All products in the supermarket were designed referring to real brands and packages. Through speakers real supermarket sounds were played as typical acoustic background stimulation. The VR supermarket was

Table 1
Patient sample characteristics.

	OctaVis-condition (n = 21)	Desktop-condition (n = 17)	*p*	*r²*	*statistic* +	*df*
gender (female)	n = 8 (38,1%)	n = 11 (64,7%)	0.191	0.045	#	1
age (years)	47.48 (±9.40)	42.29 (±13.72)	0.196	0.048	−1.33	27.3
years of education	13.52 (±2.21)	13.06 (±2.49)	0.545	0.010	−0.61	36
weekly computer usage[1]	3.95 (±2.25)	4.71 (±1,72)	0.263	0.034	1.14	36
Beck Depression Inventory (BDI-II)	28.05 (±12.29)	32.35 (±7.60)	0.216	0.041	1.26	36
symptoms: acut/remitted	71%/29%	94%/6%	0.082	0.050	#	1
antidepressant (at least one):	95%	94%	0.701	0.001	#	1
general cognitive abilities (LPS)[2]						
- general verbal knowledge (ST 1, 2)	47.95 (±7.09)	51.82 (±7.44)	0.110	0.067	1.64	36
- non-verbal inference (ST 4)	54.93 (±5.95)	56.88 (±5.62)	0.309	0.027	1.03	36
- word recognition/vocabulary (ST 12)	52.90 (±8.34)	55.76 (±6.77)	0.261	0.033	1.14	36
general immersion tendency (ITQ)[3]	49.38 (±15.03)	46.47 (±11.77)	0.518	0.011	−0.65	36

Note: [1] = rating scale: 0 = "never" to 6 = "several times a day"; [2] = given in T-scores; [3] = assessing the level to which someone is generally likely to become involved or immersed into a virtual environment; ST: subtest, ITQ: Immersive tendencies questionnaire, LPS: Leistungsprüfungssystem; r^2 = effect size measure (see Ferguson, 2009); + = unpaired *t*-test, # = Exact Fisher Test.

modeled according to a real standard medium-sized (25 × 25 m) German supermarket. Several photographs of the OctaVis-device can be found in the supplemental online material.

In a former pilot study (see Dyck, Schmidt, Piefke, & Botsch, 2012), 27 university students (age range 18–35 years) were asked to move through the VR supermarket, visit specific landmarks and accomplish several tasks there (e.g. counting offers). In one group the VR supermarket was presented within the OctaVis-device and a normal single-screen notebook was used in the other group. Subjective experiences of presence while completing the tasks within the supermarket were examined by an established instrument of Witmer and Singer (1998). In conclusion, ratings of presence were significantly higher in the group of OctaVis-users (M = 130.5 ± 17.4) compared to the notebook-group (M = 104.3 ± 15.2, Wilcoxon test: $p < 0.01$). This emphasized that the VR-presentation within the OctaVis increases the degree of immersion relative to a 3D presentation of a standard notebook monitor.

In a further study, evaluating the level of immersion of the OctaVis-device and the resulting experience of presence, 19 healthy elderly participants (age range 32–94 years) and 10 stroke patients (age range 34–79) performed the above described training paradigm in the VR supermarket (Dyck et al., 2012). Ratings on a questionnaire based on the instrument of Witmer and Singer (1998) revealed that the OctaVis-device was again found to be a very immersive training environment, which induces senses of presence in the virtual supermarket situation (Dyck et al., 2012).

Results of an other evaluation study indicated that exercising within the OctaVis-environment compared to setups without surround view or rotating body motions leads to benefits in spatial orientation in people with no prior experience in VR (Dyck, Pfeiffer, & Botsch, 2013).

2.3. Comparison condition

To contrast our advanced immersive VR-device with a typical PC-desktop condition, in a comparison setting the same digital supermarket training scenario was applied, but presented just on a single 26″ LCD-touch-screen. A joystick is used for navigation and product selection is done by tapping the target object on the touch screen. Here, the reality-like 360°-all around visibility and the rotating body motions were omitted, making this setting less immersive and closely resembling to known computerized cognitive training setups.

2.4. Cognitive training program

The cognitive training program for both conditions took place on eight days (see Fig. 1) and was based on established learning paradigms, e.g. the California Verbal Learning Test (Delis, Kramer, Kaplan, & Thompkins, 1987) or the Auditory Verbal Learning Test (Schmidt, 1996). On each training trial from day 1–6 participants had to memorize the same auditorily presented target shopping list ("list A") including 20 items (*learning trials*). These items were compiled out of four semantic categories: beverages (e.g. lemonade), groceries (e.g. apples), household goods (e.g. pencils) and hygiene items (e.g. toothpaste). Subsequently, patients first had to verbally recall and afterwards buy all recognized products in the digital supermarket. On day 7, items of a distractor list ("list B") had to be recalled and purchased, followed by a reproduction and virtual shopping of the original target list A without a renewed auditory presentation (*recall after interference*). Finally, on day 8, participants had to recall and buy the target items of list A without another listening to the list (*recall after delay*). The interference and delay trial were included to consider the stability of the newly acquired learning content under usual everyday-like conditions (cf. de Wall, Wilson, & Baddeley, 1994).

Although there were no time restrictions, participants were encouraged to accomplish the task as fast as possible. While during training trials 1–5, the training intervals were between 1 and 3 days, trials 6–8 took place on consecutive days. The maximum training duration was set to 14 days.

2.5. Assessment

2.5.1. Baseline measures

All participants had been subjected to comprehensive diagnostic evaluation prior to the training procedure. Demographic (age, gender, computer use, etc.) and relevant clinical data (e.g. previous diseases, etc.) were inquired in a pre-structured interview. To control for general cognitive abilities, we examined the patients' general verbal knowledge (subtest 1, 2), non-verbal reasoning (subtest 4) and word recognition/vocabulary (subtest 12) using a well-established German intelligence test (LPS, Leistungsprüfungssystem Horn, 1983). Moreover, to check for participants' tendency to get involved or immersed into a virtual environment, the Immersive Tendencies Questionnaire was performed (ITQ, Witmer & Singer, 1998) before training.

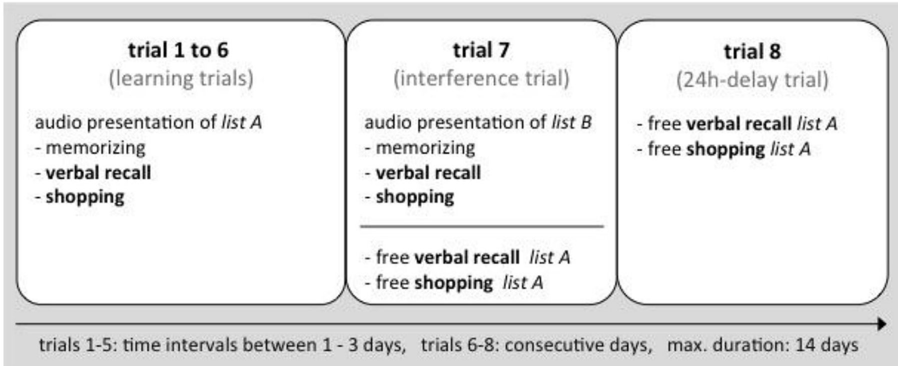

Fig. 1. Overview of the VR training procedure (*During six trials of learning, patients had first to memorize and then to verbally recall and to buy items of an auditorily presented target list (list A). On trial 7, at first an interfering list (list B) was presented and patients had to recall and buy items of this list B. Secondly, patients had to recall and buy itmes of the target list A, but without another presentation of list A. On the last trial, patients had to recall and buy only items of the target list A again without any further presentation*).

2.5.2. Psychopathological measures

Depressive symptoms were assessed using the Structured Clinical Interview for DSM-IV (SKID, Wittchen, Zaudig, & Fydrich, 1997), which was conducted by trained psychologists. The clinical examination also included assessment of the severity of depressive symptoms (Beck Depression Inventory, BDI-II, Hautzinger, Keller, & Kühner, 2006).

2.5.3. Cognitive complaint measures

Evaluations of patients' cognitive complaints in everyday life were captured by the Questionnaire for Complaints of Cognitive Disturbances (Fragebogen zur geistigen Leistungsfähigkeit, FLei, Beblo, Kunz et al., 2010). It covers the areas of attention (e.g. “While reading a novel I lose track of action and people.”), memory (e.g. “If I read a newspaper article, I forget its content.”) and executive functions (e.g. “It is difficult for me to plan my day.”).

2.5.4. Neuropsychological measures

Alertness as well as selective attention were assessed by subtests of the computer-based test battery for attention performance (TAP, Zimmermann & Fimm, 2009). Regarding learning and memory we assessed (a) verbal short-term memory and (b) verbal working memory (Digit Span Task (forward and backward) of the Wechsler Memory Scale – Revised, WMS-R; Härting et al., 2000), (c) verbal learning and memory (Verbaler Lern-und Merkfähigkeitstest, VLMT, German version of the Rey Auditory Verbal Learning Test; Helmstaedter, Lendt, & Lux, 2001) as well as (d) visuo-spatial memory (Rey/Taylor-Complex-Figure, RCFT, Meyers & Meyers, 1995). Concerning executive functions, patients processed semantic and lexical verbal fluency tasks (Regensburg Verbal Fluency Test, Regensburger Wortfluessigkeitstest, RWT; Aschenbrenner, Tucha, & Lange, 2000). We also included two measures of visuo-spatial perception with low (Bergen Right-Left Discrimination Test, BRLD; Ofte & Hughdahl, 2002) and high demands on mental rotation abilities (BRLD; Grewe, Ohmann, Markowitsch, & Piefke, 2014).

As part of the pre-post-evaluation, questionnaires and tests were administered again after training-completion, if available, using parallel-forms.

2.6. Real-life shopping task

In a medium-sized local supermarket, which was unknown to all patients, real-life shopping performance was assessed similar to the computerized training procedure: At first, patients were given a sheet of paper with a 20 items shopping list (unlike the training list, but items based on the same categories), which they had to read aloud once. Then they were given 2 min to learn the items before lastly reading it out once again. Subsequently, patients had to verbally recall all remembered products. Afterwards, they start to “purchase” them in the supermarket. Purchasing a remembered product was operationalized by tipping on the respective product and tell the name aloud. Walking distance was assessed by a trajectory counting wheel which was installed on the shopping cart. The total shopping time was acquired with a standard stopwatch.

2.7. Performance measures

As measures of verbal learning and memory performance both the number of correctly recalled products of the target list prior to shopping as well as the number of correctly purchased products in the VR-supermarket were considered for statistical analysis. In order to assess improvements in processing speed and spatial navigation, shopping time and walking distance in the VR supermarket were also analyzed.

In the real-life shopping task, the number of correctly purchased products from the previously obtained list, walking distance as well as shopping time were considered for analysis.

2.8. Data analysis

All statistical analyses were conducted using SPSS 20 (SPSS Inc., Chicago, Illinois). The general significance level was set to $\alpha = 0.05$ (two-tailed) if not otherwise specified. Basis demographic and clinical data were analyzed by t-tests or exact Fisher tests.

Group-differences and progress of the performances across the entire training procedure was examined through repeated-measures ANOVAs, considering the following trials (targeting “list A”): the very first trial as baseline (trial 1), the last training trial (trial 6) and the trials after interference (trial 7) and 24-h delay (trial 8). The shift of performance after interference and delay compared to the last learning trial was of particular interest. So if applicable, post-hoc tests were conducted using paired t-tests.

The hypothesized advantage of the advanced VR-environment was tested by the interaction of the factors “group” (OcatVis versus desktop) and “time” (pre-versus post-training) in two-way repeated-measures analyses of variance (ANOVA) using a) cognitive complaints, b) real-shopping performance and c) neuropsychological variables as dependent outcome variables. Here, significant interaction effects (group x time) were expected in favor of the novel VR-environment.

If the assumption of normal distribution was violated as indicated by Kolmogorov-Smirnov-Test, suitable non-parametric tests were performed. In the case of multiple testing (e.g. post-hoc-tests), a corrected significance level was considered according to the Bonferroni method (α divided by number of conducted comparisons). Appropriate effect sizes (η_p^2, r^2)[1] were calculated (Ferguson, 2009).

In addition to the results presented here, more detailed descriptive statistical data involving means and standard deviations of all the outcome measures can be obtained from the supplemental online material.

3. Results

3.1. Training performances

3.1.1. Verbal recall and purchased products

As outlined in Fig. 2, verbal memory and learning performance increased in the course of the training-trials in both groups. Repeated-measures ANOVA of trials 1, 6, 7 and 8 supported the effect of time on the number of *verbally recalled products* (F(2,43; 84,99) = 161.01, $p < 0.001$, $\eta_p^2 = 0.821$), but showed no group-differences (F(1; 35) = 0.459, $p = 0.502$, $\eta_p^2 = 0.013$) or an interaction-effect (F(2,43; 84,99) = 3.13, $p = 0.399$, $\eta_p^2 = 0.027$). This was also the case for the *correctly purchased products* in the VR-supermarket as Friedman tests showed effects of time for both groups (Desktop: $\chi^2(16) = 30.32$, $p < 0.001$, $r^2 = 0.948$; OctaVis: $\chi^2(21) = 47.76$, $p < 0.001$, $r^2 = 1.00$), while no between-group difference appeared at any of the four time-points (trial 1: T(36) = −0.813, $p = 0.422$, $r^2 = 0.022$; trial 6: U = 142.00, $p = 0.264$, $r^2 = 0.035$; trial 7: U = 169.50, $p = 0.799$, $r^2 = 0.002$; trial 8: T(35) = −0.343, $p = 0.733$, $r^2 = 0.003$).

Taking into account the corrected significance level (α/6 = 0.0083), post-hoc tests showed a significant gain in the *correctly recalled products* from the first to the last learning trial for both groups (desktop: $T(16) = -16.11$, $p < 0.001$, $r^2 = 0.791$; OctaVis: Z = −5.39, $p < 0.001$, $r^2 = 0.692$). After interference and delay, there was no significant decrease in the correctly recalled products neither in the desktop-condition (interference: $T(16) = 1.23$, $p = 0.236$, $r^2 = 0.011$; delay: $T(15) = 0.141$, $p = 0.889$, $r^2 < 0.001$) nor in the OctaVis-condition (interference: $T(20) = 1.90$, $p = 0.072$, $r^2 = 0.024$; delay: $T(20) = -0.181$, $p = 0.858$, $r^2 < 0.001$). The number of *correctly purchased products* had risen across the learning trials in both conditions (Desktop: $T(16) = -13.28$, $p < 0.001$, $r^2 = 0.713$; OctaVis: $Z = -4.02$, $p < 0.001$, $r^2 = 0.385$). The amount of products did not change after interference or delay both in the desktop- (interference: $T(15) = 2.10$, $p = 0.053$, $r^2 = 0.047$; delay: $T(15) = 0.735$, $p = 0.474$, $r^2 = 0.006$) and the OctaVis-group (interference: $T(20) = 1.23$, $p = 0.234$, $r^2 = 0.004$; delay: $T(20) = 2.35$, $p = 0.030$, $r^2 = 0.017$).

3.1.2. Shopping time and distance

Shopping time and walking distance decreased across the training-trials in both conditions (see Fig. 3). Throughout the trials 1, 6, 7 and 8 there was a significant effect of time on required shopping time (repeated-measures ANOVA: $F(1.75, 59.56) = 59.67$, $p < 0.001$, $\eta_p^2 = 0.637$), but no interaction-effect ($F(1.75, 59.56) = 0.369$, $p = 0.665$, $\eta_p^2 = 0.011$) or any group-differences ($F(1, 34) = 0.080$, $p = 0.779$, $\eta_p^2 = 0.002$). The same pattern occurred for the results of the walked distance (time: $F(1.62,55.11) = 36.75$,

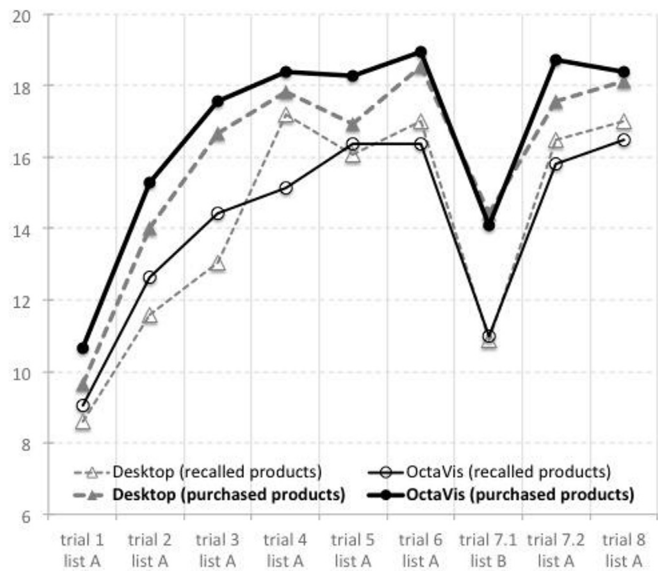

Fig. 2. Different measures of memory performance across the VR-training trials for both conditions (list A = target learning list, list B = interference/distractor list).

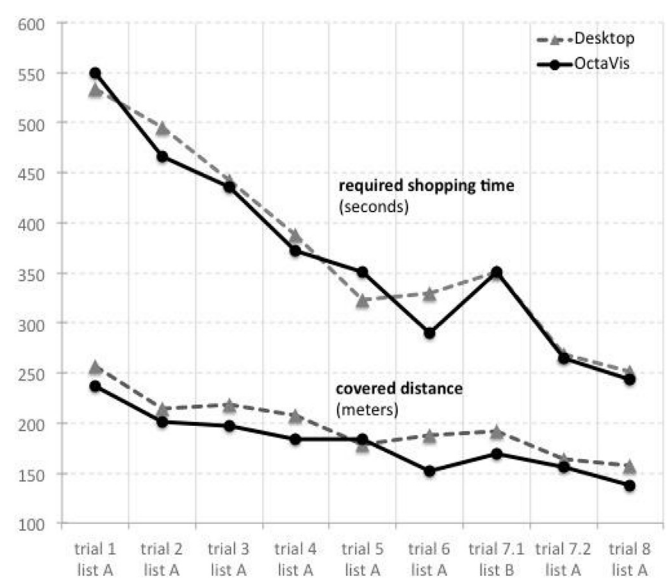

Fig. 3. Measures of speed and navigation performance across the VR-training trials for both conditions (list A = target learning list, list B = interference/distractor list).

$p < 0.001$, $\eta_p^2 = 0.519$; interaction: $F(1.62, 55.11) = 0.494$, $p = 0.574$, $\eta_p^2 = 0.014$; group: $F(1,34) = 1.51$, $p = 0.228$, $\eta_p^2 = 0.042$).

Considering the corrected post-hoc significance level (corrected $\alpha = 0.0083$), post-hoc tests showed a significant improvement in the time needed for shopping from the first to the last learning trial for both groups (desktop: $T(15) = 6.01$, $p < 0.001$, $r^2 = 0.271$; OctaVis: $T(20) = 5.26$, $p < 0.001$, $r^2 = 0.313$). In the desktop-condition there was no significant decrease of the shopping time in the trials after interference ($T(15) = 2.44$, $p = 0.028$, $r^2 = 0.047$) and after 24hour-delay ($T(15) = 2.76$, $p = 0.015$, $r^2 = 0.087$) compared to the last learning trial. In the OctaVis-group, the mean required shopping time remained stable after interference ($T(20) = 1.20$, $p = 0.246$, $r^2 = 0.012$) and after the 24 h-delay-trial ($T(20) = 2.80$, $p = 0.011$, $r^2 = 0.038$). The *walking distance* dropped from trial 1 to trial 6 in both groups (desktop: $T(15) = 3.92$,

[1] Effect size interpretation suggestions by Ferguson (2009): minimum effect size = 0.04, moderate effect size = 0.25, strong effect size = 0.64.

$p = 0.001$, $r^2 = 0.130$; OctaVis: $T(20) = 4.64$, $p < 0.001$, $r^2 = 0.256$). There was no significant decrease of the walked distance after interference and delay in the desktop- (interference: $T(15) = 1.84$, $p = 0.086$, $r^2 = 0.046$; delay: $T(15) = 1.94$, $p = 0.071$, $r^2 = 0.066$) as well as in the OctaVis-condition (interference: $T(20) = -0.484$, $p = 0.634$, $r^2 = 0.003$; delay: $T(20) = 2.29$, $p = 0.033$, $r^2 = 0.046$).

3.1.3. Relationship between training performances and depressive symptoms

To examine possible associations between improvements of learning and depressive symptoms (BDI-II), difference scores were calculated and correlated with each other (e.g. purchased products at trial 1 subtracted from bought products at trial 6). There was a significant decrease of depressive symptoms (BDI-II) across the study procedure ($F(1,36) = 27.32$, $p < 0.001$, $\eta_p^2 = 0.431$), but no difference between groups ($F(1,36) = 2.76$, $p = 0.106$, $\eta_p^2 = 0.071$) or an interaction effect ($F(1, 36) = 0.225$, $p = 0.638$, $\eta_p^2 = 0.006$). Mean BDI-II-scores significantly reduced as following: From 32.35(±7.60) to 24.29(±11.04) in the desktop-group and from 28.05(±12.29) to 18.38(±11.07) in the OctaVis-group. Neither in the desktop- ($p = 0.165$) nor in the OctaVis-condition ($p = 0.331$) any significant correlation of depressive symptoms with differences scores of verbal recall, purchased products, walked distance or shopping time was observed (see Table A5 in the supplemental material). Thus, progresses in learning performances appeared to be statistically independent of the reduction in depression symptom severity.

3.2. Cognitive complaints

Both training groups reported strong cognitive complaints in everyday-life: Prior to training, they exceeded the mean of a healthy subjects sample by nearly two standard deviations (n = 97; memory: 10.79(±6.60); attention: 9.67(±6.53); executive functions: 8.65(±6.67); Beblo, Kunz et al., 2010). Impairments were still high after the training, ranging between one and two standard deviations above the means of the healthy control subjects of the Beblo et al. study.

The two groups were comparable regarding cognitive complaints prior to (but also after) the training, as indicated by non-significant group-effects (see Table 2). Furthermore, regarding the pre-post changes in subjective impairments, non of the repeated measures ANOVAs revealed a significant interaction (group x time) effect in favor of the OctaVis-condition.

Only for exploration, we compared both groups with respect to changes within the three cognitive domains, because results of ANOVAs yielded significant main effects of time for cognitive impairments in all domains (Table 2). Here, exploratory post-hoc t-tests (corrected $\alpha = 0.025$) showed significant reductions in impairments-ratings for the OctaVis-condition only (attention: $T(20) = 2.94$, $p = 0.008$, $r^2 = 0.033$; memory: $T = 2.77$, $p = 0.012$, $r^2 = 0.025$; executive functions: $T = 2.67$, $p = 0.015$, $r^2 = 0.033$), but not in the desktop group (attention: $T(16) = 1.12$, $p = 0.279$, $r^2 = 0.022$; memory: $T = 2.22$, $p = 0.042$, $r^2 = 0.032$; executive functions: $T = 0.436$, $p = 0.668$, $r^2 = 0.002$). In addition to the fact that these are solely explorative analyzes, it should be noted that all effect sizes are very small (η_p^2, $r^2 \leq 0.033$) and did not reach the proposed minimum level of 0.04 (Ferguson, 2009).

3.3. Real-life shopping task

As shown in Table 2, there were no pre-post performance changes after the training, no differences between both groups and also no interaction effect for a) the verbal recall of products prior to shopping, b) the required shopping time or c) the distance walked inside the real supermarket (see Table 2). Regarding the correctly purchased products, no main time- or group-effect, but a significant interaction effect occurred. This was based on an increase of correctly bought products from pre- ($M = 12.18 \pm 3.08$) to post-training performance ($M = 14.13 \pm 2.75$) in the desktop-condition

Table 2
Repeated-measures analyses for the subjective impairment ratings and the real-life shopping task.

	repeated measures ANOVA: (T)ime, (G)roup or (I)nteraction (group x time) effects			
	T	I	G	df
subjective cognitive impairments				
attention	$F = 7.05$, $p = 0.012$, $\eta_p^2 = 0.164$	$F = 0.549$, $p = 0.463$, $\eta_p^2 = 0.015$	$F = 0.009$, $p = 0.926$, $\eta_p^2 < 0.001$	1, 36
memory	$F = 12.08$, $\mathbf{p = 0.001}$, $\eta_p^2 = 0.251$	$F = 0.236$, $p = 0.630$, $\eta_p^2 = 0.007$	$F = 0.361$, $p = 0.552$, $\eta_p^2 = 0.010$	1, 36
executive functions	$F = 4.22$, $\mathbf{p = 0.047}$, $\eta_p^2 = 0.105$	$F = 1.89$, $p = 0.178$, $\eta_p^2 = 0.050$	$F = 0.015$, $p = 0.902$, $\eta_p^2 < 0.001$	1, 36
real-life shopping task				
correctly recalled products prior to shopping	$F = 1.96$, $p = 0.173$, $\eta_p^2 = 0.063$	$F = 1.65$, $p = 0.209$, $\eta_p^2 = 0.054$	$F = 4.01$, $p = 0.055$, $\eta_p^2 = 0.121$	1, 29
correctly bought products	$F = 1.39$, $p = 0.248$, $\eta_p^2 = 0.044$	$F = 4.59$, $\mathbf{p = 0.040}$, $\eta_p^2 = 0.133$	$F = 2.43$, $p = 0.130$, $\eta_p^2 = 0.075$	1,30
required shopping time	$F = 3.72$, $p = 0.063$, $\eta_p^2 = 0.110$	$F = 1.76$, $p = 0.195$, $\eta_p^2 = 0.055$	$F = 0.018$, $p = 0.894$, $\eta_p^2 = 0.001$	1, 30
walked distance inside the supermarket	$F = 0.454$, $p = 0.506$, $\eta_p^2 = 0.017$	$F = 0.263$, $p = 0.613$, $\eta_p^2 = 0.010$	$F = 2.26$, $p = 0.145$, $\eta_p^2 = 0.080$	1, 26

df = degrees of freedom; see supplemental material for detailed descriptive statistics. Statistically significant differences (p < 0.05) are marked bold.

in contrast to a slight decrease in the OctaVis-group (M_{pre} = 12.13 ± 2.96, M_{post} = 11.56 ± 2.78). As a limitation, it has to be stated, that with respect to the corrected significance level ($\alpha = 0.025$), just a statistical trend did arise here for the increased desktop-group performance in the post-hoc-analysis ($T(15) = -2.26$, $p = 0.039$, $r^2 = 0.099$; OctaVis: $T(15) = 0.72$, $p = 0.488$, $r^2 = 0.010$).

3.4. Neuropsychological measures

The two groups were comparable regarding cognitive performances in all but one neuropsychological variables prior to the training procedures. Results of the repeated-measures ANOVAs showed not any significant group differences – apart from verbal working-memory (see Table 3). Here, post-hoc t-tests (corrected $\alpha = 0.025$) indicated higher scores of verbal working-memory for the OctaVis-group both before ($T(36) = -2.42$, $p = 0.021$, $r^2 = 0.135$) and after the training procedure ($T(36) = -3.06$, $p = 0.004$, $r^2 = 0.200$).

Across all other neuropsychological variables, analyses revealed no single significant interaction (time x group) effect in favor of the more real-life-like OctaVis-condition (see Table 3).

As a significant effect of time on cognitive performance appeared for phasic alertness, visuo-spatial memory measures, and visuo-spatial imagery (see Table 3), we exploratorily compared both groups with respect to pre-post-changes in these measures. Post-hoc paired t-tests (corrected $\alpha = 0.025$) indicated that only desktop-groups phasic alertness reaction significantly increased ($T(16) = -2.56$, $p = 0.021$, $r^2 = 0.047$) between assessments (OctaVis: $T(17) = -1.30$, $p = 0.213$, $r^2 = 0.014$). Post-hoc tests for the 3-min visuo-spatial recall revealed a significant pre-post-enhancement in the OctaVis-group ($T(20) = -3.46$, $p = 0.002$, $r^2 = 0.065$) as well as a statistical trend in the desktop-condition ($T(16) = -2.34$, $p = 0.032$, $r^2 = 0.076$). The performance in the 30-min recall significantly improved in both the OctaVis- ($T(20) = -2.44$, $p = 0.024$, $r^2 = 0.043$) and the desktop-group ($T(16) = -3.35$, $p = 0.004$, $r^2 = 0.122$). Visuo-spatial imagery only increased significantly in the OctaVis-group both in the low

Table 3
Repeated-measures analyses for neuropsychological variables.

	repeated-measures ANOVA: (T)ime, (I)nteraction (group x time) or (G)roup effects			
	T	I	G	df
alertness (without signal)	$F = 0.009$, $p = 0.925$, $\eta_p^2 < 0.001$	$F = 0.303$, $p = 0.586$, $\eta_p^2 = 0.009$	$F = 0.057$, $p = 0.813$, $\eta_p^2 = 0.002$	1, 33
alertness (with signal)	$F = 0.624$, $p = 0.435$, $\eta_p^2 = 0.019$	$F = 0.142$, $p = 0.709$, $\eta_p^2 = 0.004$	$F = 0.308$, $p = 0.583$, $\eta_p^2 = 0.009$	1, 33
phasic alertness	$F = 6{,}56$, $\boldsymbol{p = 0.015}$, $\eta_p^2 = 0.166$	$F = 0.270$, $p = 0.607$, $\eta_p^2 = 0.008$	$F = 0.001$, $p = 0.981$, $\eta_p^2 < 0.001$	1, 33
selective attention (Go/Nogo)	$F = 0.432$, $p = 0.516$, $\eta_p^2 = 0.013$	$F = 3.96$, $p = 0.055$, $\eta_p^2 = 0.107$	$F = 1{,}20$, $p = 0.282$, $\eta_p^2 = 0.035$	1, 33
Wordfluency semantic condition	$F = 1.92$, $p = 0.174$, $\eta_p^2 = 0.051$	$F = 0.3{,}09$, $p = 0.087$, $\eta_p^2 = 0.079$	$F = 0.054$, $p = 0.817$, $\eta_p^2 = 0.002$	1,36
Wordfluency semantic-switching	$F = 3{,}94$, $p = 0.055$, $\eta_p^2 = 0.099$	$F = 0.003$, $p = 0.953$, $\eta_p^2 = 0.000$	$F = 0.005$, $p = 0.943$, $\eta_p^2 = 0.000$	1, 36
Wordfluency phonematic condition	$F = 2{,}88$, $p = 0.098$, $\eta_p^2 = 0.074$	$F = 0.044$, $p = 0.835$, $\eta_p^2 = 0.001$	$F = 0.970$, $p = 0.331$, $\eta_p^2 = 0.026$	1, 36
Wordfluency phonematic-switching	$F = 0.009$, $p = 0.925$, $\eta_p^2 = 0.000$	$F = 0.303$, $p = 0.586$, $\eta_p^2 = 0.009$	$F = 0.057$, $p = 0.813$, $\eta_p^2 = 0.002$	1, 33
verbal short-term memory (digit-span forward)	$F = 0.753$, $p = 0.391$, $\eta_p^2 = 0.020$	$F = 0.090$, $p = 0.766$, $\eta_p^2 = 0.002$	$F = 2.57$, $p = 0.117$, $\eta_p^2 = 0.067$	1, 36
verbal working memory (digit-span backward)	$F = 2{,}09$, $p = 0.157$, $\eta_p^2 = 0.055$	$F = 0.673$, $p = 0.417$, $\eta_p^2 = 0.018$	$F = 10{,}15$, $\boldsymbol{p = 0.003}$, $\eta_p^2 = 0.220$	1, 36
visuo-spatial memory (3 min recall)	$F = 13{,}98$, $\boldsymbol{p = 0.001}$, $\eta_p^2 = 0.280$	$F = 0.836$, $p = 0.367$, $\eta_p^2 = 0.023$	$F = 0.375$, $p = 0.544$, $\eta_p^2 = 0.010$	1, 36
visuo-spatial memory (30 min recall)	$F = 18{,}69$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.342$	$F = 3{,}69$, $p = 0.063$ $\eta_p^2 = 0.093$	$F = 0.428$, $p = 0.517$, $\eta_p^2 = 0.012$	1, 36
visuo-spatial imagery/mental rotation (low demands)	$F = 16.75$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.318$	$F = 1.95$, $p = 0.171$ $\eta_p^2 = 0.051$	$F = 0.998$, $p = 0.325$, $\eta_p^2 = 0.027$	1, 36
visuo-spatial imagery/mental rotation (high demands)	$F = 14.55$, $\boldsymbol{p = 0.001}$, $\eta_p^2 = 0.288$	$F = 0.492$, $p = 0.488$ $\eta_p^2 = 0.013$	$F = 1.60$, $p = 0.214$, $\eta_p^2 = 0.043$	1, 36

df = degrees of freedom; see supplemental material for detailed descriptive statistics.

demand ($T(20) = -5.38$, $p < 0.001$, $r^2 = 0.111$) and the high demand condition ($T(20) = -4.12$, $p = 0.001$, $r^2 = 0.099$). Improvements in the desktop-group did not reach the corrected significance level (low demands: $T(16) = -1.47$, $p = 0.161$, $r^2 = 0.023$, high demands: $T(16) = -1.76$, $p = 0.097$, $r^2 = 0.024$).

4. Discussion

In this study, we examined the efficacy of a real-life-like VR-environment aiming to improve ecological validity in cognitive rehabilitation for patients with depression in comparison to a desktop-VR application. Since previous computer-based cognitive training approaches had been criticized for insufficient assessments of transfer effects (e.g. Jak et al., 2013), we used a comprehensive examination of self-ratings of everyday-life related cognitive impairment, a real-life task as well as neuropsychological tests.

The main outcome of the present study is that patients with depression did not show a greater benefit of the cognitive training in an immersive everyday simulating environment compared to a desktop application. First of all, there was no advantage in transfer effects regarding everyday-life related cognitive complaints, real life shopping task outcomes, and different neuropsychological domains (except working memory). Moreover, performances of both groups were comparable across all examined training measures (recognition, performance speed, spatial orientation). The present results might be explained and discussed in the light of a) the questionable effect of the additional degree of immersion, b) the patient's situational and motivational factors influencing the ecological validity as well as c) methodological and statistical constraints.

4.1. *Effect of the degree of immersion*

First of all, our results are in line with other studies not supporting superior effects of more immersive virtual environments as opposed to standard desktop screens. For example, in a non-clinical sample (n = 42, age range 14–40), Santos et al. (2009) aimed to examine whether the use of a fully immersive VR system (head-mounted display, HMD) would bring up greater navigational benefits than using a traditional desktop setup. Results of the applied game-like navigational task showed that some outcomes were comparable between the two conditions and the global user performance (e.g. collected items, walked distance, average speed) was even significantly better for the desktop setup. This has been ascribed to the fact that the typical desktop condition presents a well known configuration, while most of the subjects had never used a fully immersive VR system. A similar explanation was given by Felnhofer, Heinzle, and Kothgassner (2013), who did not find differences in user ratings of presence and immersion after performing a 3D computer role playing game either on an immersive head-mounted display (HMD) or a standard flat screen monitor. Furthermore, participants (52 healthy women, age range 19–36) reported stronger experiences of flow in the flat screen condition and found it to be easier to use and to access. These results had been interpreted as indicating that the requirements were probably "*more demanding*" using the HMD as opposed to "*the more commonly used and thus, more familiar TV-flat screen*" (Felnhofer et al., 2013, p. 5). In our study, subjects were familiarized with the OctaVis device through detailed instructions and a practice trial (using a virtual office room) prior to the training. Nevertheless, it cannot be ruled out that adapting to the training condition within the completely novel VR system was perhaps more demanding than adapting to the rather common desktop application. This might explain why the more immersive setup of the OctaVis-device in relation to the PC desktop did not lead to improved transfer effects of the cognitive training. With regard to clinical practice, it should be considered that, depending on the patient group, an increase in immersion does not seem to be necessarily associated with an enhanced effectiveness of VR (see also Tortella-Feliu et al., 2011). Existing less-immersive computer-assisted cognitive trainings have also been found to improve cognitive performance (Grynszpan et al., 2011; Bowie, Porter, Bowie, Jordan, & Malhi, 2013).

However, some other studies comparing desktop screens with other highly immersive setups have shown that a greater degree of immersion could result in an increase of task performance, for instance in tasks of procedure memorization (Ragan, Sowndararajan, Kopper, & Bowman, 2010) or psychomotor training (Stevens & Kincaid, 2015). Admittedly, it should be noted that differing results might be attributable to varying participant samples, study purposes as well as applied task settings.

4.1.1. *Differential training effects*

Although patients did not show a superior performance due to OctaVis-training compared to a single-screen desktop application, exploratory pre-post-analysis indicated that there are perhaps some differential effects of the training conditions on separate neuropsychological capacities: Whereas enhancements in visuospatial memory appeared in both groups, an improvement of phasic alertness was found for the desktop-condition only and mental rotation performance solely improved in the OctaVis-group. It has to be pointed out, that the corresponding effect sizes altogether showed rather modest to small effects for the differences of the training method ($r^2 = 0.043–0.122$). Moreover, the findings of these exploratory analyses have to be considered as solely indicative of differential training effects.

It is not entirely clear whether improvements in alertness are due to the desktop training requirements (e.g. staying visually concentrated and focused on just one desktop-screen vs. complex interaction within the 360°-VR-environment) or if it is grounded in a high training-assessment-correlation, since attention-testing itself is done computerized using a laptop-screen. Contrary to our results, Cho et al. (2004) found, that a neurofeedback training with a highly immersive HMD-based setup was more effective in enhancing attention-performance than a low-immersive VR condition, which only used a computer monitor with a fixed viewpoint. Restrictively, it must be noted, that the sample consisted of children with deficits of inattention and impulsiveness, rather than depressive symptoms.

The finding of improved mental rotation performance after the OctaVis-training is in line with other studies, where VR-training has also been found to enhance visual-spatial abilities in different clinical samples (Kober et al., 2013; Passig & Eden, 2001; Rose et al., 2005). Requirements of the OctaVis-training, e.g. coordinating the navigation by using 360°-rotating body-movements in close balancing with joystick locomotion-control (vs. using just the joystick to navigate in the desktop-condition) are quite likely accountable for significant improvements in visual-spatial imagery and mental rotation abilities. It was shown elsewhere that allowing VR users to control simulated rotations with their own body can have benefits over joystick control in a navigational search task (Dyck et al., 2013; Riecke et al., 2010).

4.2. *Ecological validity of the training setting*

Missing superiority of the OctaVis-condition compared to a desktop application might be due to situational and motivational factors concerning the dissimilarity of training context and everyday life. It may be assumed that the OctaVis training was still too much embedded within the emulated, experimental setting of the underlying study situation. Some differences between mostly artificial assessment situations and patient's real-world environments have already been pointed out to address the general relevance of ecological validity in neuropsychology (Chaytor & Schmitter-Edgecombe, 2003; Sbordone, 2008): Assessments typically take place in quiet and defined settings, widely free of distracting or highly emotional stimuli, based on circumscribed timeframes, the examiner offers one-to-one instructions, initiates the tasks execution, monitors its progress and takes up a supportive as well as forbearing behavior. By contrast, everyday life is rather unstructured, complex, partly noisy or stressful, enriched with individual emotional items and requiring the patient to act self-reliantly. Applied to the current study, considerable differences like detailed instructions, standardization of the procedure, a precisely defined task (to buy 20 predefined products) and a reasonable length of time also occurred here in contrast to daily demands of the patients (e.g. familiarizing oneself with a demand, deciding independently or organizing autonomously). It has been suggested, that in persons with depression a facilitation of structuredness is associated with decreased rumination and simultaneously enhanced memory functions (see Gotlib & Joormann, 2010).

Moreover, the underlying situation or purpose of an investigation itself appears to be an important influence factor: In contrast to the most common daily situations, psychological assessments in general are particularly suited to engender an atmosphere of feeling challenged or being tested. The whole current study situation features some relevant characteristics as well, like the unusual real-life-shopping *task* or the comprehensive *diagnostic* session prior to training. Not to mention the notion of participating in a study-project with all its formalities and standardizations. A large Europe-wide study ($n = 763$) aiming to examine attitudes to psychiatric research of patients with schizophrenia and depression revealed that about half of the participants stated that taking part in research would make them feel "tested" (Schäfer et al., 2011). The impression and perception of "feeling tested" or the participation in research by itself might then impact the participants behavior compared to the otherwise unobserved situation, a construct often described as the "Hawthorne effect" (McCambridge, Witton, & Elbourne, 2014).

To some extent such differences between "unobserved" everyday functioning and "investigated" cognitive performance of the patients with depression finally also appeared in the present study. It could be seen that the patients on average complain about marked cognitive impairments in their everyday life, while they showed nondescript performances across the training process. For instance, even though patients in both groups stated strong impairments in their daily memory functions – which exceeded a healthy control group's mean by nearly two standard deviations – they, however, achieved significantly interference- and delay-resistant results in recognition performance (verbally recalled and purchased products). Furthermore, although depressive patients in both groups indicate marked cognitive impairments in their everyday life, they achieved normal results in the neuropsychological examination (less than one SD below the normative mean; see Table A2 in the supplemental material). In neuropsychological assessments this would typically be interpreted as "unimpaired" (see Lezak, 2004). The disparity between subjective and objective cognitive performance of patients with depression found in our study coincide with results of former investigations. Besides a provable under-estimation of memory-functions in depressed persons (Biringer, Sundet, & Lund, 2009), both mixed psychiatric inpatient samples involving depressed patients (Moritz, Ferahli, & Naber, 2004) as well as several studies using distinct groups of patients with depression (Beblo, Mensebach et al., 2010; Lahr et al., 2007; Mowla et al., 2007) yielded absent relationships between subjective and objective measures of neurocognition. This was recently also confirmed in a German large-scale epidemiological investigation (Wolfsgruber et al., 2015). Whether this disparity is solely based on the above mentioned differences between assessment situation and everyday performance or is maybe more influenced by negative-biased self-perceptions originating from symptomatic negative cognitive patterns appearing in depression (Disner, Beevers, Haigh, & Beck, 2011) needs to be further explored.

4.3. Limitations and perspectives

Some constraints have to be considered. First of all, it should be noted that this study is based on a non-random and small patient sample. Our patients were assigned to groups in a non-random convenience assignment depending on availability for the testing and training sessions during their inpatient-treatment. However, both groups were comparable a) according to sociodemographic, clinical, and VR-related variables prior to training and b) no differences in baseline training, real-life or neuropsychological measures were found (except for verbal working memory), which suggested that there were no relevant group differences.

According to the sample size, our study is comparable to other studies investigating the effectiveness or comparability of computerized cognitive trainings in psychiatric patients (Alvarez, Sotres, León, Estrella, & Sosa, 2008) or healthy participants (e.g. Smith, Stibric, & Smithson, 2013; Whitlock, McLaughlin, & Allaire, 2012). Nevertheless, in small-scaled studies significance testing might be problematic in general: Smaller sample sizes are associated with smaller power, which means that the chance of discovering existing effects is low and risk of false-negative results is increased (Button et al., 2013). Therefore, we examined if the absence of significant interaction effects may be due to low power and conducted a post-hoc power analyses using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). Here it is assumed, that the comprehensive reality-oriented VR-environment with its elaborate hardware setup should at least have a moderate effect to have an incremental benefit over a typical PC desktop application. This power analysis revealed that, for detecting a moderate-sized population effect ($f = 0.25$) at a two-tailed alpha-level of 0.05, the effective power in our study ($n = 38$, 2 groups, 4 measurements) was 0.96. If one makes use of a slightly smaller effect ($f = 0.20$), then the determined power was still 0.83. A power of > 0.80 is usually considered as a minimum requirement (Cohen, 1992), what caused us to assume that our results can not be explained as being due to low statistical power. Furthermore, the relevance of significant results is generally indicated by effect size measures, since the effect size is hardly influenced by the sample size (Ferguson, 2009). In our small-scaled study, the interaction effect sizes (η_p^2=0.001–0.107) are almost all smaller or only slightly larger than the aforementioned minimum effect size suggestion of 0.04 (Ferguson, 2009), regardless of the non-significant p-value results. This again supports our main finding, that there is no (additional) effect of cognitive training in an immersive everyday simulating environment compared to a desktop application in patients with depression.

As a methodological limitation, it should be mentioned, that the digital simulated supermarket with its moderate size, manageable complexity and missing distractors (e.g. no other present persons) may provide a task too easy for the participants, lacking some real daily requirements. Furthermore, the whole training paradigm used in our study comprises a relatively short period of time only (range: 8–14 days). In a meta-analysis of cognitive remediation programs for schizo-affective and affective disorders a median duration of intervention of 12 weeks was found, although frequency and duration of training sessions varied considerably between the included studies (Anaya et al., 2012). In respect thereof, our cognitive training program could be seen far from sufficient and increasing the number of trials or frequency of training should be considered in further studies. In this context, variations in the presented learning material as well as in the simulated surrounding context should also be implemented to possibly further increase generalization effects of the training. While we used the real-shopping task as a functional performance measure, a future study could compare the application of the VR-based cognitive training program with a corresponding "real" supermarket training intervention. The comparison of both conditions might be helpful to gain more precise information on requirements for an ecologically valid VR-based training program.

Our sample consisted of patients both in acute and remitted states of major depressive disorder. Neurocognitive impairments have been found to generally improve with remission of depression in most of the patients (Beblo et al., 2011). Thus, although additional explorative analyses regarding the acute/remitted-status did not reveal any statistical differences in our results, it might be interesting to explicitly examine in further studies, whether the course of depression has an impact on the effect of (VR-based) cognitive training.

4.4. Conclusion

In sum, the current results did not prove an advantage of a comprehensive virtual reality environment for cognitive rehabilitation of depressed patients in comparison to a typical desktop setup. Since ecological validity constrains in neuropsychological assessments has been repeatedly emphasized, the present findings also imply difficulties in improving transfer effects of computerized cognitive rehabilitation using virtual reality. Thereby, the current study gives rise to future research developing other approaches to improve ecological validity of neuropsychological therapy for patients with depression.

Acknowledgements

This work was supported by the project CITmed (Cognitive Interaction Technology in Medicine), funded by the EFRE (Europäischer Fonds für Regionale Entwicklung) program of Northrhine Westfalia, and the Cluster of Excellence Cognitive Interaction Technology (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.chb.2017.10.019.

References

Alvarez, L. M., Sotres, J. F. C., León, S. O., Estrella, J., & Sosa, J. J. S. (2008). Computer program in the treatment for major depression and cognitive impairment in university students. *Computers in Human Behavior, 24*(3), 816–826. https://doi.org/10.1016/j.chb.2007.02.013.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Arlington: American Psychiatric Publishing.

Anaya, C., Martinez Aran, A., Ayuso-Mateos, J. L., Wykes, T., Vieta, E., & Scott, J. (2012). A systematic review of cognitive remediation for schizo-affective and affective disorders. *Journal of Affective Disorders, 142*(1–3), 13–21. https://doi.org/10.1016/j.jad.2012.04.020.

Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *Regensburger Wortflüssigkeits-test (RWT)*. Bern: Huber.

Beblo, T., Kunz, M., Brokate, B., Scheurich, A., Weber, B., Albert, A., et al. (2010). Entwicklung eines Fragebogens zur subjektiven Einschätzung der geistigen Leistungsfähigkeit (FLei) bei Patienten mit psychischen Störungen. *Zeitschrift für Neuropsychologie, 21*(3), 143–151. https://doi.org/10.1024/1016-264X/a000013.

Beblo, T., Mensebach, C., Wingenfeld, K., Schlosser, N., Rullkoetter, N., Schaffrath, C., et al. (2010). The impact of neutral and emotionally negative distraction on memory performance and its relation to memory complaints in major depression. *Psychiatry Research, 178*(1), 106–111.

Beblo, T., Sinnamon, G., & Baune, B. T. (2011). Specifying the neuropsychology of affective Disorders: Clinical, demographic and neurobiological factors. *Neuropsychology Review, 21*(4), 337–359. https://doi.org/10.1007/s11065-011-9171-0.

Biringer, E., Sundet, K., & Lund, A. (2009). P01-229 Self-evaluation of memory functioning in depressed patients during a symptomatic episode and after recovery. *European Psychiatry, 24*, S617. https://doi.org/10.1016/S0924-9338(09)70850-8.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience, 12*, 752–762. https://doi.org/10.1038/nrn3122.

Bora, E., Harrison, B. J., Yücel, M., & Pantelis, C. (2013). Cognitive impairment in euthymic major depressive disorder: A meta-analysis. *Psychological Medicine, 43*(10), 2017–2026. https://doi.org/10.1017/S0033291712002085.

Bortolato, B., Carvalho, A., & McIntyre, R. (2015). Cognitive dysfunction in major depressive disorder: A state-of-the-art clinical review. *CNS & Neurological Disorders - Drug Targets, 13*(10), 1804–1818. https://doi.org/10.2174/1871527313666141130203823.

Bossert, M., Weisbrod, M., & Aschenbrenner, S. (2014). Kognitive Remediation im klinischen Alltag: Eine Studie zur Akzeptanz bei psychiatrischen Patienten. *Fortschritte der Neurologie, Psychiatrie, 82*(12), 691–694. https://doi.org/10.1055/s-0034-1385477.

Bowie, C. R., Gupta, M., & Holshausen, K. (2013). Cognitive remediation therapy for mood Disorders: Rationale, early evidence, and future directions. *The Canadian Journal of Psychiatry, 58*(6), 319–325.

Bryck, R. L., & Fisher, P. A. (2012). Training the brain: Practical applications of neural plasticity from the intersection of cognitive neuroscience, developmental psychology, and prevention science. *American Psychologist, 67*(2), 87–100. https://doi.org/10.1037/a0024657.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. https://doi.org/10.1038/nrn3475.

Chan, C. L., Ngai, E. K., Leung, P. K., & Wong, S. (2010). Effect of the adapted virtual reality cognitive training program among Chinese older adults with chronic schizophrenia: A pilot study. *International Journal of Geriatric Psychiatry, 25*(6), 643–649. https://doi.org/10.1002/gps.2403.

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181–197.

Choi, J., & Medalia, A. (2005). Factors associated with a positive response to cognitive remediation in a community psychiatric sample. *Psychiatric Services, 56*(5), 602–604.

Cho, B. H., Kim, S., Shin, D. I., Lee, J. H., Min Lee, S., Young Kim, I., et al. (2004). Neurofeedback training with virtual reality for inattention and impulsiveness. *Cyberpsychology & Behavior, 7*(5), 519–526.

Conradi, H. J., Ormel, J., & De Jonge, P. (2011). Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study. *Psychological Medicine, 41*(6), 1165–1174.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Cornwell, B. R., Salvadore, G., Colon-Rosario, V., Latov, D. R., Holroyd, T., Carver, F. W., … Grillon, C. (2010). Abnormal hippocampal functioning and impaired spatial navigation in depressed individuals: Evidence from whole-head magnetoencephalography. *American Journal of Psychiatry, 167*(7), 836–844. https://doi.org/10.1176/appi.ajp.2009.09050614.

Coyle, H., Traynor, V., & Solowij, N. (2015). Computerized and virtual reality cognitive training for individuals at high risk of cognitive Decline: Systematic review of the literature. *The American Journal of Geriatric Psychiatry, 23*(4), 335–359. https://doi.org/10.1016/j.jagp.2014.04.009.

Cummings, J. J., & Bailenson, J. N. (2015). How immersive is Enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740.

Delis, D. C., Kramer, J. H., Kaplan, E., & Thompkins, B. A. O. (1987). *California verbal learning test (CVLT): Adult version: Manual*. Psychological Corporation.

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology, 6*, 26. https://doi.org/10.3389/fpsyg.2015.00026.

Disner, S. G., Beevers, C. G., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*(8), 467–477. https://doi.org/10.1038/nrn3027.

Dyck, E., Pfeiffer, T., & Botsch, M. (2013). Evaluation of surround-view and self-rotation in the OctaVis VR-System. In *Proceedings of the 5th joint virtual reality conference* (pp. 1–8). Eurographics Association.

Dyck, E., Schmidt, H., Piefke, M., & Botsch, M. (2012). OCTAVIS: Optimization techniques for multi-GPU multi-view rendering. *Journal of Virtual Reality and Broadcasting, 9*(6).

Dyck, E., Zell, E., Kohsik, A., Grewe, P., Winter, Y., Piefke, M., et al. (2012). OCTAVIS: An easy-to-use VR-system for clinical studies. In J. Bender, A. Kuijper,

D. W. Fellner, & Guérin (Eds.), *Workshop on virtual reality interaction and physical simulation (VRIPHYS)*. The Eurographics Association.

Evans, V. C., Iverson, G. L., Yatham, L. N., & Lam, R. W. (2014). The relationship between neurocognitive and psychosocial functioning in major depressive disorder: A systematic review. *Journal of Clinical Psychiatry, 75*(12), 1359–1370.

Falconer, C. J., Slater, M., Rovira, A., King, J. A., Gilbert, P., Antley, A., et al. (2014). Embodying compassion: A virtual reality paradigm for overcoming excessive self-criticism. *PLoS One, 9*(11), e111933. https://doi.org/10.1371/journal.pone.0111933.

Fals-Stewart, W., & Lam, W. K. K. (2010). Computer-assisted cognitive rehabilitation for the treatment of patients with substance use disorders: A randomized clinical trial. *Experimental and Clinical Psychopharmacology, 18*(1), 87–98. https://doi.org/10.1037/a0018058.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Felnhofer, A., Heinzle, A.-K., & Kothgassner, O. D. (2013). Game experience and behavior in young women: A comparison of interface technologies. In *Proceedings of the usability day XI, Dornbirn/Austria*.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538. https://doi.org/10.1037/a0015808.

Foster, P. P. (2015). Role of physical and mental training in brain network configuration. *Frontiers in Aging Neuroscience, 7*. https://doi.org/10.3389/fnagi.2015.00117.

Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When presence and emotion are related, and when they are not. In *8th annual international workshop on presence, September* (pp. 21–23).

Gates, N. J., Sachdev, P. S., Singh, M. A. F., & Valenzuela, M. (2011). Cognitive and memory training in adults at risk of dementia: A systematic review. *BMC Geriatrics, 11*(1), 55.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: The virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking, 14*(3), 99–105.

Gotlib, I. H., & Joormann, J. (2010). Cognition and Depression: Current status and future directions. *Annual Review of Clinical Psychology, 6*(1), 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305.

Gould, N. F., Holmes, M. K., Fantie, B. D., Luckenbaugh, D. A., Pine, D. S., Gould, T. D., et al. (2007). Performance on a virtual reality spatial memory navigation task in depressed patients. *American Journal of Psychiatry, 164*(3), 516–519.

Grewe, P., Kohsik, A., Flentge, D., Dyck, E., Bien, C., Winter, Y., et al. (2013). Learning real-life cognitive abilities in a novel 360-virtual reality supermarket: A neuropsychological study of healthy participants and patients with epilepsy. *Journal of Neuroengineering and Rehabilitation, 10*(1), 42.

Grewe, P., Ohmann, H. A., Markowitsch, H. J., & Piefke, M. (2014). The Bergen left–right discrimination test: Practice effects, reliable change indices, and strategic performance in the standard and alternate form with inverted stimuli. *Cognitive Processing, 15*(2), 159–172. https://doi.org/10.1007/s10339-013-0587-8.

Grynszpan, O., Perbal, S., Pelissolo, A., Fossati, P., Jouvent, R., Dubal, S., et al. (2011). Efficacy and specificity of computer-assisted cognitive remediation in schizophrenia: A meta-analytical study. *Psychological Medicine, 41*(1), 163–173. https://doi.org/10.1017/S0033291710000607.

Hampstead, B. M., Gillis, M. M., & Stringer, A. Y. (2014). Cognitive rehabilitation of memory for mild cognitive impairment: A methodological review and model for future research. *Journal of the International Neuropsychological Society, 20*(2), 135–151. https://doi.org/10.1017/S1355617713001306.

Härting, C., Markowitsch, H. J., Neufeld, H., Calabrese, P., Deisinger, K., & Kessler, J. (2000). *WMS-R Wechsler Gedächtnistest - Revidierte Fassung der Wechsler Memory Scale*. Bern: Huber.

Harvey, P. D. (2011). Mood symptoms, cognition, and everyday functioning: In major depression, bipolar disorder, and schizophrenia. *Innovations in Clinical Neuroscience, 8*(10), 14.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck depressions Inventar (BDI-II)*. Frankfurt: Harcourt Test Services.

Helmstaedter, C., Lendt, M., & Lux, S. (2001). *Verbaler Lern- und Merkfähigkeitstest (VLMT)*. Weinheim: Beltz-Test.

Horn, W. (1983). *Leistungsprüfungssystem (LPS) [manual] (2)*. Göttingen: Hogrefe.

Hviid, L. B., Ravnkilde, B., Ahdidan, J., Rosenberg, R., Stødkilde-Jørgensen, H., & Videbech, P. (2010). Hippocampal visuospatial function and volume in remitted depressed patients: An 8-year follow-up study. *Journal of Affective Disorders, 125*(1), 177–183. https://doi.org/10.1016/j.jad.2010.01.002.

Jak, A. J., Seelye, A. M., & Jurick, S. M. (2013). Crosswords to computers: A critical review of popular approaches to cognitive enhancement. *Neuropsychology Review, 23*(1), 13–26. https://doi.org/10.1007/s11065-013-9226-5.

Juan, M. C., & Pérez, D. (2009). Comparison of the levels of presence and anxiety in an acrophobic environment viewed via HMD or CAVE. *Presence, 18*(3), 232–248.

Keilp, J. G., Gorlyn, M., Russell, M., Oquendo, M. A., Burke, A. K., Harkavy-Friedman, J., et al. (2013). Neuropsychological function and suicidal behavior: Attention control, memory and executive dysfunction in suicide attempt. *Psychological Medicine, 43*(3), 539–551. https://doi.org/10.1017/S0033291712001419.

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research, 51*(1), S225–S239.

Knight, R. G., & Titov, N. (2009). Use of virtual reality tasks to assess prospective memory: Applicability and evidence. *Brain Impairment, 10*(01), 3–13.

Kober, S. E., Wood, G., Hofer, D., Kreuzig, W., Kiefer, M., & Neuper, C. (2013). Virtual reality in neurologic rehabilitation of spatial disorientation. *Journal of Neuroengineering and Rehabilitation, 10*(1), 1.

Kurtz, M. M., Moberg, P. J., Gur, R. C., & Gur, R. E. (2001). Approaches to cognitive remediation of neuropsychological deficits in schizophrenia: A review and meta-analysis. *Neuropsychology Review, 11*(4), 197–210.

Lahr, D., Beblo, T., & Hartje, W. (2007). Cognitive performance and subjective complaints before and after remission of major depression. *Cognitive Neuropsychiatry, 12*(1), 25–45.

Laver, K. E., George, S., Thomas, S., Deutsch, J. E., & Crotty, M. (2015). Virtual reality for stroke rehabilitation. In The Cochrane Collaboration (Ed.), *Cochrane database of systematic reviews*. Chichester, UK: John Wiley & Sons, Ltd.

Lee, R. S. C., Hermens, D. F., Naismith, S. L., Lagopoulos, J., Jones, A., Scott, J., et al. (2015). Neuropsychological and functional outcomes in recent-onset major depression, bipolar disorder and schizophrenia-spectrum disorders: A longitudinal cohort study. *Translational Psychiatry, 5*(4), e555. https://doi.org/10.1038/tp.2015.50.

Lee, R. S. C., Hermens, D. F., Porter, M. A., & Redoblado-Hodge, M. A. (2012). A meta-analysis of cognitive deficits in first-episode major depressive disorder. *Journal of Affective Disorders, 140*(2), 113–124. https://doi.org/10.1016/j.jad.2011.10.023.

Leung, I. H., Walton, C. C., Hallock, H., Lewis, S. J., Valenzuela, M., & Lampit, A. (2015). Cognitive training in Parkinson disease - a systematic review and meta-analysis. *Neurology, 85*(21), 1843–1851.

Lezak, M. D. (2004). *Neuropsychological assessment*. USA: Oxford University Press.

Lindenmayer, J.-P., McGurk, S. R., Mueser, K. T., Khan, A., Wance, D., Hoffman, L., et al. (2008). A randomized controlled trial of cognitive remediation among inpatients with persistent mental illness. *Psychiatric Services, 59*(3), 241–247.

Lopresti, E. F., Mihailidis, A., & Kirsch, N. (2004). Assistive technology for cognitive rehabilitation: State of the art. *Neuropsychological Rehabilitation, 14*(1–2), 5–39. https://doi.org/10.1080/09602010343000101.

Marques, A., Queirós, C., & Rocha, N. (2008, September). Virtual reality and neuropsychology: A cognitive rehabilitation approach for people with psychiatric disabilities. In *ICDVRAT–proceedings of 7th international conference on disability virtual reality and associated technologies (Sep 8-11 2008). Maia, Portugal* (pp. 39–46).

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015.

McIntyre, R. S., Cha, D. S., Soczynska, J. K., Woldeyohannes, H. O., Gallaugher, L. A., Kudlow, P., … Baskaran, A. (2013). Cognitive deficits and functional outcomes in major depressive disorder: Determinants, substrates, and treatment interventions. *Depression and Anxiety, 30*(6), 515–527.

Meyers, J. E., & Meyers, K. R. (1995). *Rey complex figure test and recognition trial (RCFT)*. Lutz, FL: Psychological Assessment Resources.

Moritz, S., Ferahli, S., & Naber, D. (2004). Memory and attention performance in psychaitric patients. *Journal of the International Neuropsychological Society, 10*, 623–633.

Mowla, A., Ashkani, H., Ghanizadeh, A., Dehbozorgi, G. R., Sabayan, B., & Chohedri, A. H. (2007). Do memory complaints represent impaired memory performance in patients with major depressive disorder? *Depression and Anxiety, 25*(10), E92–E96. https://doi.org/10.1002/da.20343.

Ofte, S. H., & Hughdahl, K. (2002). Right-left discrimination in male and female, young and old subjects. *Journal of Clinical and Experimental Neuropsychology, 24*(1), 81–92.

Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., … Bordin, A. (2010). Controlling memory impairment in elderly adults using virtual reality memory training: A randomized controlled pilot study. *Neurorehabilitation and Neural Repair, 24*(4), 348–357. https://doi.org/10.1177/1545968309353328.

Papakostas, G. I. (2014). Cognitive symptoms in patients with major depressive disorder and their implications for clinical practice. *The Journal of Clinical Psychiatry, 75*(1), 8–14. https://doi.org/10.4088/JCP.13r08710.

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *The Clinical Neuropsychologist, 27*(8), 1328–1361. https://doi.org/10.1080/13854046.2013.834971.

Parsons, T. D. (2011). Neuropsychological assessment using virtual environments: Enhanced assessment technology for improved ecological validity. In S. Brahnam, & L. C. Jain (Eds.), *Advanced computational intelligence paradigms in healthcare* (Vol. 6, pp. 271–289). Berlin: Springer.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00660.

Passig, D., & Eden, S. (2001). Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children. *CyberPsychology & Behavior, 4*(6), 681–686.

Porter, R. J., Bowie, C. R., Jordan, J., & Malhi, G. S. (2013). Cognitive remediation as a treatment for major depression: A rationale, review of evidence and recommendations for future research. *Australian and New Zealand Journal of Psychiatry, 47*(12), 1165–1175.

Ragan, E. D., Sowndararajan, A., Kopper, R., & Bowman, D. A. (2010). The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments, 19*(6), 527–543.

Richard-Devantoy, S., Jollant, F., Kefi, Z., Turecki, G., Olié, J. P., Annweiler, C., et al. (2012). Deficit of cognitive inhibition in depressed elderly: A neurocognitive marker of suicidal risk. *Journal of Affective Disorders, 140*(2), 193–199. https://doi.org/10.1016/j.jad.2012.03.006.

Riecke, B. E., Bodenheimer, B., McNamara, T. P., Williams, B., Peng, P., & Feuereissen, D. (2010). Do we need to walk for effective virtual reality navigation? Physical rotations alone may suffice. In C. Hölscher, T. F. Shipley, M. O. Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *International conference spatial cognition 2010 Proceedings (spatial cognition VII)* (pp. 234–247). Berlin/Heidelberg: Springer.

Riva, G., Mantovani, F., & Gaggioli, A. (2004). Presence and rehabilitation: Toward second-generation virtual reality applications in neuropsychology. *Journal of Neuroengineering and Rehabilitation, 1*(1), 1.

Rizzo, A. A., Buckwalter, J. G., Neumann, U., Kesselman, C., & Thiebaux, M. (1998). Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities. *CyberPsychology & Behavior, 1*(1), 59–78.

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation, 14*(1–2), 207–239.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine, 44*(10), 2029–2040. https://doi.org/10.1017/S0033291713002535.

Rose, F. D., Brooks, B. M., & Rizzo, A. A. (2005). Virtual reality in brain damage rehabilitation: Review. *CyberPsychology & Behavior, 8*(3), 241–262.

Santos, S. B., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., et al. (2009). Head-mounted display versus desktop for 3D navigation in virtual reality: A user study. *Multimedia Tools and Applications, 41*(1), 161–181. https://doi.org/10.1007/s11042-008-0223-2.

Sbordone, R. J. (2008). Ecological validity of neuropsychological testing: Critical issues. *The Neuropsychology Handbook*, 367–394.

Schäfer, I., Burns, T., Fleischhacker, W. W., Galderisi, S., Rybakowski, J. K., Libiger, J., et al. (2011). Attitudes of patients with schizophrenia and depression to psychiatric research: A study in seven european countries. *Social Psychiatry and Psychiatric Epidemiology, 46*, 159–165.

Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook*. Los Angeles: Western Psychological Services.

Schultheis, M. T., Himelstein, J., & Rizzo, A. A. (2002). Virtual reality and neuropsychology: Upgrading the current tools. *The Journal of Head Trauma Rehabilitation, 17*(5), 378–394.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006.

Shin, H., & Kim, K. (2015). Virtual reality for cognitive rehabilitation after brain injury: A systematic review. *Journal of Physical Therapy Science, 27*(9), 2999.

Smith, S. P., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior, 29*(6), 2388–2393. https://doi.org/10.1016/j.chb.2013.05.014.

Stevens, J. A., & Kincaid, J. P. (2015). The relationship between presence and performance in virtual simulation training. *Open Journal of Modelling and Simulation, 3*(2), 41–48. https://doi.org/10.4236/ojmsi.2015.32005.

Tarr, M. J., & Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience, 5*, 1089–1092. https://doi.org/10.1038/nn948.

Tchanturia, K., Davies, H., Lopez, C., Schmidt, U., Treasure, J., & Wykes, T. (2008). Letter to the Editor: Neuropsychological task performance before and after cognitive remediation in anorexia nervosa: A pilot case-series. *Psychological Medicine, 38*(9), 1371–1373. https://doi.org/10.1017/S0033291708003796.

Teo, W. P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., … Ayaz, H. (2016). Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation?–A narrative review of the literature. *Frontiers in Human Neuroscience, 10*, 284. https://doi.org/10.3389/fnhum.2016.00284.

Tortella-Feliu, M., Botella, C., Llabrés, J., Bretón-López, J. M., del Amo, A. R., Baños, R. M., et al. (2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behavior Modification, 35*(1), 3–30.

Vocci, F. J. (2008). Cognitive remediation in the treatment of stimulant abuse disorders: A research agenda. *Experimental and Clinical Psychopharmacology, 16*(6), 484–497. https://doi.org/10.1037/a0014101.

de Wall, C., Wilson, B. A., & Baddeley, A. D. (1994). The extended rivermead behavioural memory test: A measure of everyday memory performance in normal adults. *Memory, 2*(2), 149–166. https://doi.org/10.1080/09658219408258942.

Weiss, P. L., Kizony, R., Feintuch, U., & Katz, N. (2006). Virtual reality in neurorehabilitation. *Textbook of Neural Repair and Rehabilitation, 51*(8), 182–197.

Whitlock, L. A., McLaughlin, A. C., & Allaire, J. C. (2012). Individual differences in response to cognitive training: Using a multi-modal, attentionally demanding game-based intervention for older adults. *Computers in Human Behavior, 28*(4), 1091–1096. https://doi.org/10.1016/j.chb.2012.01.012.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.

Wittchen, H.-U., Zaudig, M., & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV (SKID)*. Göttingen: Hogrefe.

Wolfsgruber, S., Wagner, M., Kleineidam, L., Maier, W., Wittchen, U., Jacobi, F., et al. (2015). Das Verhältnis von subjektiven und objektiven kognitiven Einschränkungen bei depressiven Störungen. In *Presentation (HS-26, 003) at Kongress der Deutschen Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde (DGPPN), Berlin*.

Wykes, T., Huddy, V., Cellard, C., McGurk, S. R., & Czobor, P. (2014). A meta-analysis of cognitive remediation for schizophrenia: Methodology and effect sizes. *American Journal of Psychiatry, 168*, 472–485.

Zimmermann, P., & Fimm, B. (2009). *Testbatterie zur Aufmerksamkeitsprüfung (TAP) Version 2.2 [manual]*. Herzogenrath: Psytest.