Does cognitive flexibility training enhance subjective mental functioning in healthy older adults?

Buitenweg, J.I.V.; Van De Ven, R.M.; Ridderinkhof, K.R.; Murre, J.M.J.

DOI
10.1080/13825585.2018.1519106

Publication date
2019

Document Version
Final published version

Published in
Aging, Neuropsychology, and Cognition

License
CC BY-NC-ND

Citation for published version (APA):
Does cognitive flexibility training enhance subjective mental functioning in healthy older adults?

Jessika I. V. Buitenweg, Renate M. Van De Ven, K. Richard Ridderinkhof & Jaap M. J. Murre

To cite this article: Jessika I. V. Buitenweg, Renate M. Van De Ven, K. Richard Ridderinkhof & Jaap M. J. Murre (2019) Does cognitive flexibility training enhance subjective mental functioning in healthy older adults?, Aging, Neuropsychology, and Cognition, 26:5, 688-710, DOI: 10.1080/13825585.2018.1519106

To link to this article: https://doi.org/10.1080/13825585.2018.1519106

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 17 Sep 2018.

Article views: 610

View Crossmark data
Does cognitive flexibility training enhance subjective mental functioning in healthy older adults?

Jessika I. V. Buitenweg, Renate M. Van De Ven, K. Richard Ridderinkhof and Jaap M. J. Murre

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands; Amsterdam Brain & Cognition (ABC), University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT
Declining cognitive abilities in older adults can contribute to significant changes in socioemotional health and substantially reduce their perception of well-being. Whereas much attention has been dedicated to creating cognitive training programs to improve cognitive health in old age, there is little emphasis on the consequences of such interventions for subjective mental functioning. We created a randomized controlled trial in which we evaluated the effects of an adaptive computerized cognitive flexibility training. Healthy older adults (60–80 years old) were assigned to one of three conditions (frequent or infrequent switching or active control group) and performed 58 half-hour sessions within a period of 12 weeks. We measured effects on subjective cognitive failures and executive dysfunctioning, everyday functioning, depressive symptoms, anxiety, and quality of life, before, and after training. Additionally, participants’ proxies rated their cognitive failures and executive dysfunctioning. Subjective cognitive failures and executive dysfunctioning improved 4 weeks posttraining in all groups, although effect sizes were low ($\eta_p^2 = .058$ and .079, respectively) and there were no differences between groups (all $p$’s > .38). No significant changes in subjective reports were seen directly after training, which was the case in all groups. Proxies did not report any functional changes over time, yet their evaluations were significantly more favorable than those of the participants, both pretraining ($p < .0005$) and posttraining ($p = .004$). Although we found no evidence of improvement on subjective mental functioning, we adduce several factors that encourage further research into the effects of computerized cognitive training on subjective performance.

ARTICLE HISTORY
Received 3 March 2018
Accepted 29 August 2018

KEYWORDS
Aging; cognitive training; well-being; subjective mental functioning; videogames

Introduction
The world’s population is aging rapidly. According to the latest estimates, between 2015 and 2050, the proportion of adults over the age of 60 will double, yielding a disproportionate percentage of older adults (World Health Organization, 2015). Normal aging is often characterized by a decline in cognitive functions, affecting cognitive control,
processing speed, long-term memory, and other processes essential to independent daily life (Fisk & Sharp, 2004; Luo & Craik, 2008; Salthouse, 1996). Although emotional regulation and positive affect are often found to be increased in older adults (Charles & Carstensen, 2007), declining cognitive health can lead to a preoccupation with waning functional abilities, reduced social interactions, significant decreases in quality of life, and danger of depression (Cacioppo, Berntson, Bechara, Tranel, & Hawkley, 2011; Jonker, Comijs, Knipscheer, & Deeg, 2009; Pusswald et al., 2015).

In addition to objective decline in cognitive abilities, subjective cognitive complaints (Hohman, Beason-Held, Lamar, & Resnick, 2011; Waldorff, Siersma, & Waldemar, 2009), depression (Green et al., 2003; Wilson et al., 2002), and high levels of anxiety (Bierman, Comijs, Jonker, & Beekman, 2007; Sinoff & Werner, 2003) have also been noted as risk factors for subsequent cognitive decline and impairment, nursing home placement, and dementia. To assess this, subjective reports may be able to identify impending impairment at early stages, when clinical changes cannot yet be detected by neuropsychological tests (Jungwirth et al., 2008; Molinuevo et al., 2017; Schultz et al., 2015). It is, therefore, important to examine whether subjective well-being can benefit from a cognitive intervention. The present study aims to investigate the effects of computerized cognitive flexibility training on subjective cognitive, executive and everyday functioning, quality of life, depressive symptoms, and anxiety. We will refer to all these aspects throughout as subjective mental functioning.

In recent years, there has been increasing attention for the effects of cognitive training in healthy older adults (Lampit, Hallock, & Valenzuela, 2014; Melby-Lervåg, Redick, & Hulme, 2016). Although training has in many cases led to significant improvements on laboratory tasks, the subjective effects of such interventions, such as effects on quality of life or daily functioning, are often overlooked. The focus of most training studies is on objective functional effects, rather than individual experience. However, subjective mental functioning could benefit from cognitive interventions in a number of ways, for instance by directly improving self-esteem (Burgener, Yang, Gilbert, & Marsh-Yant, 2008), daily functioning (Corbett et al., 2015), and quality of life (Wolinsky et al., 2006), or diminishing depressive symptoms (Calkins, McMorran, Siegle, & Otto, 2015). Ultimately, for cognitive training effects to be meaningful and to influence older adults’ independence, it is essential for such generalization to take place (Kelly et al., 2014). Hence, an intervention should not only improve cognitive functioning, but should preferably also affect subjective cognitive experience and daily functioning in the personal environment.

Thus far, results from interventions targeting subjective outcome measures are mixed. Improvement has been found after trainings of memory enhancement (Fairchild & Scogin, 2010; Preiss, Lukavský, & Steinova, 2010; Valentijn et al., 2005), group-based self-efficacy (Hastings & West, 2009; McDougall et al., 2010), and speed of information processing (Smith et al., 2009). Likewise, leisure activities and volunteer work seem generally beneficial (Tesky, Thiel, Banzer, & Pantel, 2011), especially affecting well-being (Bureš, Čech, Mikulecká, Ponce, & Kuca, 2016; Kuykendall, Tay, & Ng, 2015; Souza, Lautert, & Hilleshein, 2011). General training of computer or Internet use, on the other hand, does not appear to have effects on quantitative subjective scales measured (Slegers, Van Boxtel, & Jolles, 2008; White et al., 2002), although participants in these training studies reported benefiting in more qualitative ways, such as in self-
confidence and social aspects. In contrast, when computer use involves playing cognitive games, participants may benefit in several domains, such as cognitive failures, social connectedness, and fewer complaints of depression (Allaire et al., 2013; Hardy et al., 2015; Hausknecht, Schell, Zhang, & Kaufman, 2015). Whether subjective experience is altered by training thus seems to depend largely on the type of intervention given, with computer game training offering promising prospects of improving subjective mental functioning.

Conversely, it is evident from these results that effectiveness is also highly dependent on the domain of subjective measurement. For instance, benefits are often seen in memory complaints (Fairchild & Scogin, 2010; Hastings & West, 2009; McDougall et al., 2010; Tesky et al., 2011), subjective attention (Richmond, Morrison, Chein, & Olson, 2011), and quality of life (Hausknecht et al., 2015; Slegers et al., 2008; Wolinsky et al., 2006). In relation to activities of daily life, some studies find no effects (Ball et al., 2002; Slegers et al., 2008), whereas others do (Corbett et al., 2015; McDougall et al., 2010). Similarly, subjective cognitive functioning has been positively influenced by training (Hardy et al., 2015; Preiss et al., 2010; Smith et al., 2009), but this has not been reported consistently (Tesky et al., 2011; Valentijn et al., 2005). Finally, improvements in mood (Fairchild & Scogin, 2010; McDougall et al., 2010; Slegers et al., 2008; Tesky et al., 2011) and self-efficacy (McDougall et al., 2010; Slegers et al., 2008) seem especially difficult to achieve.

Interpretations of subjective training results are rendered more difficult by the limited use of subjective measures in such studies, leading to possible subjective effects of most interventions not being detected. As an intervention’s test battery generally includes higher numbers of objective than subjective measurements, this might be one reason why subjective effects often go undetected (McDougall et al., 2010; Richmond et al., 2011; Shatil, Mikulecka, Bellotti, & Bureš, 2014). One important adjustment, then, would be to integrate the use of subjective measurements in future assessments of cognitive interventions. One other way to obtain a more reliable view on self-report questionnaire results is to administer the same questionnaires to a significant other or a next-of-kin family member (so-called proxies). Especially in older adults with commencing health- or cognitive problems, it could be useful to solicit the observation of a proxy (Neumann, Araki, & Gutterman, 2000).

Previous research suggests that stimulating alternating cognitive processes in subjects fosters transfer to untrained behavioral measures (Bherer et al., 2008; Buitenweg, Murre, & Ridderinkhof, 2012; Karbach & Kray, 2009). The ability to adjust one’s responses to the demand of a new situation, known as cognitive flexibility (Buitenweg et al., 2012), is a core executive function, facilitating the functioning of higher order functions (Berry et al., 2016; Buttelmann & Karbach, 2017). That being the case, cognitive flexibility proves essential for both cognitive and everyday functioning (Logue & Gould, 2014) and has been seen to be ameliorated by aerobic exercise (Masley, Roetzheim, & Gualtieri, 2009), positive mood (Hirt, Devers, & McCrea, 2008), and video game play (Colzato, Van Leeuwen, Van Den Wildenberg, & Hommel, 2010; Glass, Maddox, & Love, 2013). For this reason, we especially wanted to test whether capitalizing on cognitive flexibility within our training would result in stronger intervention effects than without this element. We therefore designed a state-of-the-art cognitive flexibility training, integrating adaptive, multimodal, novel training games and frequent sessions. In view of the importance of
subjective perception of changes in cognitive health and well-being, we included several types of subjective measures, as well as proxy versions of several questionnaires.

First and foremost, as our intervention included the important element of cognitive flexibility, we focused on assessing intervention effects on subjective cognitive functioning. Accordingly, we included subjective measures of cognitive failures as well as of executive dysfunctioning. In addition to cognitive flexibility, due to the game-based nature of our training, both speed of processing and reasoning were factors present in both experimental conditions. Since processing speed has been seen to improve health-related quality of life (Wolinsky et al., 2006) and evidence from reasoning training points to effects on everyday functioning (Willis et al., 2006), we included these specific measures of quality of life and everyday functioning. Additionally, we incorporated a measure of depression and anxiety into our battery. Although mood has previously proven difficult to improve by cognitive training, including a computerized game training (Bureš et al., 2016), to our knowledge none of these interventions included components of flexibility. Cognitive flexibility has been found to be especially impaired in individuals suffering from depression and anxiety (Gualtieri & Morgan, 2008; Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010; Murphy et al., 1999), so that including flexibility as a main ingredient in the training might have important implications for these constructs. Because a significant portion of the elderly suffer from symptoms of fatigue (Avlund, 2010), it is possible that fatigue levels in our sample are affected by training 5 days a week for 12 weeks. For this reason, a measure of fatigue was included to explore whether participants’ levels of fatigue in the experimental conditions were positively or negatively affected. Following previous literature (Floyd & Scogin, 1997; Hertzog & Pearman, 2013; Hulur, Hertzog, Pearman, & Gerstorf, 2015), we also explored the correlation between subjective and objective executive functioning, and between metamemory and objective memory and depression.

The purpose of this study was to assess whether 12 weeks of cognitive flexibility training could lead to an improvement of subjective cognitive failures and executive dysfunctioning, everyday functioning, depressive symptoms, anxiety, and quality of life. First, we expected participants in the cognitive flexibility condition to demonstrate greater improvements, compared with a non-flexibility cognitive training and an active control group, on subjective cognitive failures and executive dysfunctioning, depressive symptoms and anxiety. The non-flexibility cognitive training was expected to display larger improvements compared to active controls. Second, based on previous research on measures of quality of life and everyday functioning (Willis et al., 2006; Wolinsky et al., 2006), both cognitive training conditions were expected to improve equally, yet more so compared with the active control condition. Third, we expected the cognitive flexibility condition to maintain improvement on subjective cognitive failures and executive dysfunctioning at four weeks post-training, and report higher subjective training improvements, compared to the non-flexibility cognitive training and the active control group.

Methods

Our study involved a randomized controlled double-blind design, effects of which were assessed using an extensive battery of neuropsychological paper-and-pen tests,
questionnaires, and computer tasks. Results are distributed over multiple publications. For a full detailed description of our methods, see Buitenweg, Van De Ven, Prinssen, Murre, and Ridderinkhof (2017).

Participants

We recruited participants between the ages of 60 and 85, who were willing and cognitively able to commit to participation in a 12-week computerized training program. Participants were excluded if they had previously used the training games, had current substance abuse, severe visual impairment or colorblindness, had suffered a stroke or TIA, or scored below 26 on the Telephone Interview Cognitive Status (TICS; Brandt, Spencer, & Folstein, 1988). A schematic overview of the study design can be found in Figure 1.

The final sample included 158 participants, who were randomly assigned to one of three conditions (frequent switching, infrequent switching, and mock training; described below). We used a minimization program (Minimpy; Saghaei & Saghaei, 2011) to reduce asymmetry within the groups over the factors age, gender, education, TICS score, and computer experience. Full written informed consent was given by all participants prior to participation. The study was approved by the local Ethics Committee of the University of Amsterdam and registered under number 2012-BC-2566. All procedures were conducted in compliance with the Declaration of Helsinki, relevant laws, and institutional guidelines.

Study protocol

Upon inclusion into the project, participants were requested (though not obliged) to provide a proxy (a friend or family member) who would be willing to fill out three online questionnaires both before and after training. Participants then came to the university for neuropsychological assessment and training instructions and signed an informed consent form (time = T0). A subgroup of participants came in twice for an MRI scan as well. Results of neuropsychological tests and MRI scans are reported separately. A member of the research team provided participants with detailed training instructions, which they also received in the form of instructional videos and a booklet to view at home. Subjective experience was measured at home using several online questionnaires. The order of task administration at baseline and post-training was counterbalanced over participants. Each subject was assigned a member of the research team who offered motivation and stimulation during weekly or bi-weekly telephone calls and assisted with training-related technical and scheduling issues. Participants trained for half an hour a day, five times a week, on days and times of their choice. Log-in times were monitored, and an automatic reminder was sent if a subject did not log in for more than two days. A small online log was filled out daily before and after each training session, inquiring after participants’ motivation and fatigue. After 12 weeks of training, the training portal was closed and participants completed all questionnaires again (time = T2). A selection of questionnaires was also administered after 6 weeks of training (T1), and 4 weeks after the last training session (T3). After this last assessment point, all participants received free access to all games on the training portal.
**Intervention**

We based our three training programs on the existing website [www.braingymmer.com](http://www.braingymmer.com). Games on this site are aimed at the general population, but we modified selected games and tailored them to the needs of older participants, for example by allowing longer reaction times. The order of all games and sessions in the 12-week course was preprogrammed to provide participants with approximately 30 min of game play and to prevent participants from exclusively selecting their preferred games. Feedback after finishing the games and sessions was the same in all three groups.

---

**Figure 1.** Flow chart of the study design. FS = frequent switching; IS = infrequent switching.
Cognitive training
Based on our reading of the literature and on our experience with cognitive tasks in experimental studies with older adult participants, we selected nine games, divided over the cognitive domains working memory, reasoning, and attention, that we deemed optimal to maximize cognitive flexibility. To increase variability and flexibility, we programmed the order of the games in such a manner that two consecutive games were always from different domains. Games consisted of 20 levels of increasing difficulty. After each game, feedback on performance was given with up to three stars on the screen. To operationalize adaptiveness, participants were coaxed to a higher difficulty level each time two or three stars had been attained.

Within the cognitive training, we differentiated between two groups: frequent switching and infrequent switching. In the frequent switching condition, participants played 10 games of 3 min each, forcing them to frequently switch between different tasks and functions, thus maximizing flexibility. In the infrequent switching condition, a training session consisted of three games of 10 min each, allowing for fewer switches between domains. In the first week of training, participants in both conditions played the games for 10 min each, to sufficiently familiarize themselves with the individual games. Total time spent on each game was the same for both groups after completing 12 weeks of training.

Mock training
For the mock training (MT) we used fewer games than in the experimental conditions to reduce variability. Based on our reading of the literature, and on our experience with cognitive tasks in experimental studies with older adult participants, we selected four games that presented relatively little demand on cognitive control, yet were visually stimulating. To minimize flexibility, games in one session were played for 10 min each. Although higher levels could be played if stars were attained, we directed participants to stay on a specific level for one or two weeks, thereby reducing adaptiveness.

Outcome measures
Subjective measures
Health-related quality of life was assessed by the Short Form Health Survey (SF-36; McHorney, Ware, Lu, & Sherbourne, 1994). The SF-36 consists of 36 questions, which are divided into eight separate scales, of which the Z-scores were combined into two weighted sums with higher scores representing better quality of life. Separate scales are: Physical Functioning, Role Limitations due to Physical Problems, Bodily Pain, General Health Perceptions, Vitality, Social Functioning, Role Limitations due to Emotional Problems, and General Mental Health. These scales were summed into a physical component (PCS) and a mental component (MCS) as standardized scores for the general Dutch population (Aaronson et al., 1998). Internal consistency of this questionnaire is high, with only Social Functioning values dropping for older subgroups. Construct validity is good, though on the scales Social Functioning, General Health, and Vitality, rates drop for older adults (Sullivan, Karlsson, & Ware, 1995). Test–retest reliability of all scales is good to excellent (Andresen, Bowley, Rothenberg, Panzer, & Katz, 1996).
The Cognitive Failure Questionnaire (CFQ; Broadbent, Cooper, FitzGerald, & Parkes, 1982) was used to measure subjective cognitive functioning. This questionnaire consists of 25 items on a 5-point Likert scale and a maximum score of 100, with higher scores denoting more subjective cognitive dysfunctions. The CFQ was administered to both the participant and their proxy. Participants additionally filled out this questionnaire on the follow-up 4 weeks after training. Test–retest reliability after 21 weeks and 12 months is high (Broadbent et al., 1982) and interitem reliability is excellent (Bridger, Johnsen, & Brasher, 2013).

Subjective executive functioning was measured using the Dysexecutive Functioning Questionnaire (DEX; Burgess, Alderman, Wilson, Evans, & Emslie, 1996) included in the Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). This questionnaire includes 20 items on a 5-point Likert scale and a maximum score of 80, with higher scores representing more subjective executive dysfunctions. The DEX was administered to both the participant and their proxy. Participants additionally filled out this questionnaire on the follow-up 4 weeks after training. Internal- and retest reliability and ecological validity of this questionnaire is good (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Shinagawa et al., 2007).

Everyday functioning was assessed using the Lawton & Brody Instrumental Activities of Daily Living (IADL) scale (Lawton & Brody, 1988), consisting of eight questions on a four- to six-point scale. Scores were added up to a total, with a maximum of 22 denoting worst performance. Questions that were answered with “not applicable” or “never carried out myself in my life” were replaced with the average of the participants’ remaining items. This questionnaire was administered to both the participant and their proxy. This scale has a high retest reliability and medium validity (Graf, 2008). Sensitivity and specificity of this scale to identify individuals with cognitive impairment is medium to good (Barberger-Gateau et al., 1992).

Depression and anxiety were measured using the Hospital Anxiety Depression Scale (HADS; Zigmond & Snaith, 1983), which includes 14 items on a 4-point scale, divided over the subscales Depression and Anxiety, each with a maximum score of 21. A cut-off score of 8 and above is found to be optimal to screen for possible cases in the population for both depression and anxiety (Bjelland, Dahl, Haug, & Neckelmann, 2002). Both subscales have good retest reliability beyond 6 weeks (Herrero et al., 2003) and medium concurrent and predictive validity (Herrero et al., 2003; Lisspers, Nygren, & Söderman, 1997).

The Checklist Individual Strength-Fatigue subscale (CIS-F; Vercoulen et al., 1997) was used to assess fatigue. This scale consists of 8 items, with scores ranging from 8 to 56. A score of 35 and above is regularly used as a cut-off to denote severe fatigue, leading to high sensitivity and specificity. Internal consistency and retest reliability are also high (Worm-Smeitink et al., 2017).

To assess subjective training-induced improvement after training, we designed an exit-interview with four items (attention, reasoning, memory and overall cognition) answered on a five-point Likert scale ranging from 1 (“Certainly not improved”) to 5 (“Certainly improved”). Although this measure is not validated, it serves as a necessary tool to judge the subjective notion of improvement related to the training. The four items were added to a total score for subjective improvement with a maximum of 20.
We administered a short questionnaire online before and after each daily training session to assess motivation experienced during the training. Participants rated “Training interest before session start”, “Motivation to perform well” and “Enjoyment of today’s games” on a 7-point Likert scale. Each rating was averaged over all sessions, and the three scores were summed into a total Motivation score with a maximum of 21.

Subjective memory problems were measured with a four-question interview, in which participants were asked about memory failures, and worries and hindrance related to these problems (Van Den Kommer et al., 2014). The four items were added to a total score with a range of 0 to 12.

Training performance
We calculated game- and domain high scores, highest level attained, and gain scores between the first and last session played. To obtain game high scores, we determined the percentage of the maximum possible score per level, and added up all levels to a total per game, and all scores within each domain for domain scores. Subsequently, we computed a mean training score for all three training groups separately and transformed these to Z-scores to be able to compare mock training and experimental training groups relative to each other.

Analyses
For the first hypothesis, a series of mixed-design ANOVAs was carried out with the total scores on the CFQ and DEX and subscales of the HADS as dependent variables, with group (frequent switching, infrequent switching or mock training) as the between-participants independent variable, and time points T0 (baseline) and T2 (posttraining) as a within-participants independent variable. Two similar mixed ANOVAs were run for the second hypothesis, using the physical and mental component of the SF-36 and total score of IADL as dependent variables. For the third hypothesis, an extra mixed-design ANOVA was carried out for the DEX and CFQ on all four time points and subjective training-induced improvement on all time points from T1. All participants who completed the tasks at T2 were included in the main analyses. Additionally, a per-protocol analysis was run, including only participants who completed at least 50 sessions. As exploratory analyses, we added a repeated-measures ANOVA with CIS-F as a dependent variable, as well as correlations between subjective and objective executive functioning and between metamemory and objective memory and depression. In case of a significant effect, correlations would be computed between significant measures and age, TICS score, and workouts completed, to determine whether to add them as a covariate. Education level required nonparametric correlation analysis (Spearman’s Rho); all other measures used Pearson’s correlation coefficient. To explore the extent to which individual characteristics influenced training benefits, significantly correlated covariates were added to a repeated-measures ANCOVA of the primary and secondary measures. Outliers were detected using Grubbs’ Extreme Studentized Deviation test (Grubbs, 1950). We ran analyses with and without outliers. Reported results are without outliers, unless otherwise specified. IBM SPSS Statistics for Windows, version 22 (IBM Corp., Armonk, N.Y., USA) was used for all statistical analyses. Normality was checked using Shapiro-Wilk’s test and by evaluating skewness and kurtosis. A p-value of .05 (two-tailed
if not mentioned otherwise) was considered significant. Bonferroni corrections for multiple testing were used for all analyses. Greenhouse-Geisser corrected degrees of freedom were used whenever sphericity was violated, although for the purpose of legibility the original degrees of freedom are reported.

**Results**

Of the 158 participants who completed baseline measurements, 19 dropped out for various reasons (see Buitenweg et al., 2017). Sustainers were significantly younger than dropouts (sustainers $M = 67.77$, $SD = 5.0$; dropouts $M = 72.3$, $SD = 7.8$, $t_{[20,134]} = -2.454$, $p = 0.023$). There were no differences in education level, gender, TICS score, or training group between the final sample and dropouts (all $p$'s > .19). The final analysis was based on the 139 participants who completed both T0 and T2 (age 60–80, $M = 67.8$, 60.4% female, mean years of education 13.7). Before training, the three training groups did not differ in demographic variables or baseline subjective values (see Table 1). Mean number of sessions completed at T2 was 56, and this did not vary across groups. The frequent switch training and the mock training contained a higher number of participants than the infrequent switch condition, because participants receiving MRI scans were only assigned to either the frequent switch or mock training condition. As the IADL revealed a substantial ceiling effect, it was decided not to include this questionnaire in further analyses.

<table>
<thead>
<tr>
<th>Measure</th>
<th>FS ($n = 56$)</th>
<th>IS ($n = 33$)</th>
<th>MT ($n = 50$)</th>
<th>Sign. $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (% female)</td>
<td>64.3</td>
<td>63.6</td>
<td>54</td>
<td>.51</td>
</tr>
<tr>
<td>Age</td>
<td>67.8</td>
<td>67.9</td>
<td>67.6</td>
<td>.97</td>
</tr>
<tr>
<td>Level of education</td>
<td>5.9</td>
<td>5.8</td>
<td>5.9</td>
<td>.76</td>
</tr>
<tr>
<td>TICS Score</td>
<td>35.6</td>
<td>35.7</td>
<td>35.2</td>
<td>.51</td>
</tr>
<tr>
<td>Prior computer use</td>
<td>5.6</td>
<td>5.4</td>
<td>5.8</td>
<td>.11</td>
</tr>
<tr>
<td>SF-36 Mental</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>.45</td>
</tr>
<tr>
<td>SF-36 Physical</td>
<td>0.1</td>
<td>-0.2</td>
<td>0.2</td>
<td>.17</td>
</tr>
<tr>
<td>Cognitive Failure Questionnaire</td>
<td>31.3</td>
<td>32.4</td>
<td>28.9</td>
<td>.37</td>
</tr>
<tr>
<td>Dysexecutive Functioning Questionnaire</td>
<td>14.9</td>
<td>16.2</td>
<td>14.2</td>
<td>.60</td>
</tr>
<tr>
<td>Instrumental Activities of Daily Living</td>
<td>0.8</td>
<td>1.0</td>
<td>0.5</td>
<td>.19</td>
</tr>
<tr>
<td>HADS-Depression</td>
<td>2.0</td>
<td>2.6</td>
<td>2.1</td>
<td>.53</td>
</tr>
<tr>
<td>HADS-Anxiety</td>
<td>3.2</td>
<td>2.8</td>
<td>3.0</td>
<td>.77</td>
</tr>
<tr>
<td>CIS-Fatigue subscale</td>
<td>20</td>
<td>21.1</td>
<td>18.1</td>
<td>.48</td>
</tr>
<tr>
<td>Proxy $a$</td>
<td>24.5</td>
<td>23.5</td>
<td>24.7</td>
<td>.93</td>
</tr>
<tr>
<td>Cognitive Failure Questionnaire</td>
<td>13.3</td>
<td>10.6</td>
<td>14.0</td>
<td>.41</td>
</tr>
<tr>
<td>Dysexecutive Functioning Questionnaire</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>.83</td>
</tr>
</tbody>
</table>

Note. Lower scores represent better performance except on SF-36 and TICS; $p$ values of Gender based on $\chi^2$, all others based on ANOVA; Sign. = Significance; FS = frequent-switching experimental group; IS = infrequent-switching experimental group; MT = mock training group; Level of education based on 7-point scale (from 1 = unfinished primary school to 7 = university); TICS = Telephone Interview for Cognitive Status; SF-36 = Short Form Health Survey-36; HADS = Hospital Anxiety Depression Scale; CIS = Checklist Individual Strength; Prior computer use based on 7-point scale (from 1 = less than once a month to 7 = more than 4 h a day); $a = n_{FS} = 44$, $n_{IS} = 24$, $n_{MT} = 42$. 
**Intervention effects**

The statistics of the analyses reported below are detailed in Table 2.

For the first hypothesis, Time effects for the Cognitive Failure Questionnaire did not survive Bonferroni correction, and no modulation by Group was found. Results of the DEX and HADS Depression- and Anxiety subscales also did not reveal a significant Time effect or a Time x Group interaction. Thus, none of the groups reported any improvement regarding cognitive and executive functioning or depression. For the second hypothesis, no main effect of Time or a Time x Group interaction was found for the SF-36 Mental- and Physical subscales, suggesting none of the groups noted a gain in mental or physical health.

For the third hypothesis, in which we computed an additional mixed ANOVA for the DEX and CFQ on all four time points, more positive effects were found. Significantly different scores between time points were found on both the DEX, $F(3;372) = 7.60$, $p < .0005; \eta^2_p = .058$ (see Figure 2), and the CFQ, $F(3;378) = 10.87$, $p < .0005; \eta^2_p = .079$ (see Figure 3). *Post hoc* tests using a Bonferroni correction displayed a significant improvement on the DEX between baseline (T0) and follow-up (T3; $p = .001$) as well as between posttraining (T2) and follow-up ($p = .001$) and mid-training (T1) and follow-up ($p = .002$). On the CFQ, this effect occurred between the same time points (T0 to T3: $p < .0005$; T1 to T3: $p < .0005$; T2 to T3: $p = .020$). There was, however, no main effect of Group, demonstrating that all groups equally improved across sessions, both on the DEX, $F(6;372) = 1.143$, $p = .38$, and the CFQ, $F(3;378) = 1.044$, $p = .40$.

The same analysis was conducted on the three time points (T1, T2, and T3) of subjective training-induced improvement. Results revealed significantly different scores between time points, $F(2;258) = 6.23$, $p = .004; \eta^2_p = .046$. *Post hoc* tests determined an improvement from T1 to T2 ($p = .011$) and T1 to T3 ($p = .032$), though there was only a marginally significant interaction with Group, $F(4;258) = 2.26$, $p = .072; \eta^2_p = .034$, which was mainly caused by a much lower score on T1 by the infrequent switching group.

As only 51% of participants had filled out the subjective memory scale at baseline, we could not adequately run a repeated measures analysis on this data. Thus, in Table 2, we only report the data collected on T2, which was completed by 135 participants. No difference was found between groups in reports of memory problems, motivation, or subjective training improvement. Participants’ proxies did not report significant changes in cognitive or executive functioning over time, and there was no difference between groups. Results of these measures did not change when including outliers.

Everyone improved on the training tasks, as seen from the gain scores, although mean gain was significantly higher in frequent- and infrequent switch groups compared to the mock training, $F(2,135) = 6.69$, $p < .002$, as expected due to adaptiveness of the former. Despite this training improvement, correlations between gain score or total mean training score and change in subjective functioning were not significant (all $r$’s between −.09 and .15). A similar pattern was observed for correlations with proxy reports (all $r$’s between −.02 and .04).

**Perprotocol analyses**

Main analyses were repeated for participants who completed a minimum of 50 sessions. Eight participants were excluded (one active control participant, four participants from...
Table 2. Transfer Results

<table>
<thead>
<tr>
<th>Measure</th>
<th>Group</th>
<th>Comparison</th>
<th>Time</th>
<th>Group</th>
<th>Time * Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FS (n = 56)</td>
<td>IS (n = 33)</td>
<td>MT (n = 50)</td>
<td>Time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pretraining</td>
<td>Posttraining</td>
<td>Time</td>
<td>Group</td>
<td>Time * Group</td>
</tr>
<tr>
<td>Cognitive Failure Questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.3</td>
<td>11.7</td>
<td>28.4</td>
<td>12.4</td>
<td>32.4</td>
</tr>
<tr>
<td></td>
<td>14.9</td>
<td>9.5</td>
<td>14.9</td>
<td>10.1</td>
<td>16.2</td>
</tr>
<tr>
<td>Dysexecutive Functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 Physical</td>
<td>0.1</td>
<td>0.9</td>
<td>0.2</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>SF-36 Mental</td>
<td>2.0</td>
<td>2.1</td>
<td>1.7</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>HADS-Depression subscale</td>
<td>3.2</td>
<td>2.9</td>
<td>2.7</td>
<td>2.4</td>
<td>2.8</td>
</tr>
<tr>
<td>HADS-Anxiety subscale</td>
<td>n.a</td>
<td>9.8</td>
<td>3.2</td>
<td>n.a.</td>
<td>9.1</td>
</tr>
<tr>
<td>Memory scale (T2) a</td>
<td>n.a</td>
<td>15.3</td>
<td>3.0</td>
<td>n.a.</td>
<td>15.4</td>
</tr>
<tr>
<td>Motivation scale (T2) b</td>
<td>n.a</td>
<td>13.1</td>
<td>3.4</td>
<td>n.a.</td>
<td>13.1</td>
</tr>
<tr>
<td>Subjective improvement (T2) b</td>
<td>24.5</td>
<td>12.9</td>
<td>23.1</td>
<td>15.7</td>
<td>23.5</td>
</tr>
<tr>
<td>Proxy</td>
<td>13.3</td>
<td>10.2</td>
<td>12.2</td>
<td>10.9</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Note. Lower scores represent better performance except on SF-36; p values are based on repeated-measures ANOVA unless otherwise specified; η² = partial eta squared (effect size); FS = frequent-switching experimental group; IS = infrequent-switching experimental group; MT = mock training; SF-36 = Short Form Health Survey-36; HADS = Hospital Anxiety Depression Scale; T2 = posttraining measurement; a = analysis based on Kruskal-Wallis; b = analysis based on ANOVA; c = nFS = 44, nIS = 24, nMT = 42.
the frequent switching condition, and three from the infrequent switch condition), resulting in a final sample of 131 participants \((n_{FS} = 52, n_{IS} = 30, n_{MT} = 49)\). Participants who had completed less than 50 training sessions did not differ on subjective baseline scores or demographic variables from the rest of the sample. Per-protocol training results were similar to the main analyses (see Table S1), indicating that participants who completed at least 50 sessions did not improve more on subjective mental functioning than participants who did not.
Exploratory questions

Although neither participants nor their proxies noticed significant changes over time, proxies’ scores appeared much more favorable than the self-ratings. We performed an exploratory independent t-test on data of both time points to examine whether this difference was significant. CFQ ratings on baseline were higher (signifying more dysfunctions) in participants \((M = 30.71, SE = 1.0)\) than in proxies \((M = 24.35, SE = 1.2)\), which was significant, \(t(247) = 4.123, p < .0005, d = 0.52\). Postintervention, the same effect between proxies \((M = 23.69, SE = 1.5)\), and participants \((M = 29.01, SE = 1.1)\) appeared, \(t(225) = 2.947, p = .004, d = 0.39\). On the DEX, participants’ self-ratings on baseline were higher \((M = 14.96, SE = 0.8)\) than those of proxies \((M = 13.03, SE = 1.0)\), which was also the case after the training \((M = 14.65, SE = 0.8; M = 13.10, SE = 1.2)\) although this difference was not significant.

We explored whether participants showed altered fatigue after the training. Repeated measures analysis of the CIS-fatigue subscale showed no alteration over time, \(F(1,133) = 2.09, p = .15, \eta_p^2 = .02\), and no effect of Group, \(F(1,133) = 0.55, p = .58, \eta_p^2 = .01\).

To test whether the relationship between subjective and objective memory and depression is similar to that found in previous research (Hertzog & Pearman, 2013; Hulur et al., 2015), we computed an exploratory Pearson’s correlation coefficient between scores on our subjective memory scale, the HADS-D, and objective score on verbal long term memory (RAVLT delay; see Buitenweg et al., 2017). In accordance with the literature, we found a significant correlation between metamemory and depression \((r = .358, n = 135, p < .001)\), whereas no significant correlations arose between metamemory and objective memory score \((r = -.063)\). However, due to the low number of people in our sample reporting serious memory problems and the non-validated memory scale used, we must consider these results with care.

A similar issue concerns the question whether participants’ subjective cognitive functions correspond to their objective cognitive task performances. To explore this, we computed a Pearson’s correlation coefficient between baseline DEX and CFQ scores and primary executive functions performance (task-switching, dual tasking, updating and inhibition). Correlations between these measures were low and nonsignificant (DEX: all \(r’s\) between \(-.11\) and \(.14\); CFQ: all \(r’s\) between \(-.02\) and \(.15\)).

Discussion

In this randomized controlled study, we investigated the effects of a 12-week, computerized flexibility training on subjective cognitive and executive functioning, health-related quality of life, depression, and anxiety in older adults. Research into subjective effects of cognitive is scarce, and existing studies have rendered mixed results. Yet, as self-perceived cognitive functioning is an important prerequisite for healthy aging, we felt it essential to include these measures in our battery. No improvements were observed over time between the training groups and the active control when examined directly after training. Subjective cognitive and executive functioning increased four weeks after training, but no differential effect was seen between the intervention and control groups.
Our results suggest that including specific elements such as adaptiveness and cognitive flexibility in a computerized training is not sufficient to lead to noteworthy subjective effects immediately after training. This outcome replicates results from our stroke sample (Van DeVen et al., 2017), in which we investigated subjective effects of the TAPASS training in recovering stroke patients. In this study, time effects on the DEX and CFQ appeared in all groups, including the no-contact waiting list condition. The findings also correspond with previous studies into computerized training (Bureš et al., 2016; Slegers et al., 2008; White et al., 2002), who found null results on a range of subjective measures. Conversely, our findings contradict those of Hardy et al. (2015), who found improvement on cognitive failures following training with similar games and study length to ours. A number of details are relevant to point out. First of all, in Hardy et al. (2015) participants of all ages were recruited, with the majority of ages in the young adult range. This may very well have led to faster or larger improvements than would have been seen in a strictly older population, and makes it more difficult to compare their results directly to ours. Secondly, after playing for a minimum of 15 min each day, participants in Hardy et al. (2015) were free to continue training with any of the available games. Self-selected able or enthusiastic subjects in that study might thus have been provided with (much) additional practice, likely leading to increased internal motivation, self-control, and self-perceived cognitive improvement compared with our study, in which participants were asked to adhere to a strict session schedule.

A similar situation applies to the results of Wolinsky et al. (2006), in which benefits for quality of life were found a number of years after speed of processing training. In many of our games, speed of processing certainly played a substantial role, yet no benefit to quality of life was observed. This could be due to the fact that in our intervention, speed of processing was a covert aspect of a broader training, whereas in Wolinsky et al. (2006) it was the main focus, providing participants with explicit strategies and examples for use in daily life. Besides, both Wolinsky et al. (2006) and Hardy et al. (2015) studied a very large sample of participants, allowing small differences between groups over time to stand out more easily than in our much smaller sample. With this in mind, it is important to realize that effect sizes in Hardy et al. (2015) were small, raising the question whether the results of this study can be generalized to performance in everyday life.

Regarding the sudden increase in subjective executive and cognitive functioning four weeks after training, there are several plausible explanations. First of all, there is the possibility that participants’ reports on subsequent scales are affected by previous ones. However, we deem it unlikely that repeated measurements of this scale have caused a retest effect, as the retest reliability of both questionnaires is high (Bridger et al., 2013; Preiss et al., 2010; Shinagawa et al., 2007). Secondly, we need to consider objective improvement, which is not independent of subjective experience. In Buitenweg et al. (2017) we demonstrated cognitive improvement on a number of tasks following this training, including executive functioning. It is possible that participants only became consciously aware of general cognitive improvement until sometime after the training, therefore only reporting self-perceived changes until the post-training measurement. When asked explicitly about improvements brought on by the training, however, most subjects reported not noticing any changes in cognitive abilities. Similarly, a further increase in subjective training-induced improvement failed to appear after posttraining,
suggesting subjects’ subjective reports of cognitive and executive performance would not have been conscious.

An unexpected finding was that participants rated their own cognitive performance in terms of failures as less favorable than their proxies. Though our results are confirmed by previous research on this topic (Burgess et al., 1998), many others report individual- and proxy ratings to be equal (Bogod, Mateer, & Macdonald, 2003; Chan, 2001). However, one must keep in mind that in most studies, proxy ratings are used in comparison to patient data, with proxies of dementia patients even rating them as worse than patients themselves (Howland et al., 2017; Macháčová, Vaňková, Holmerová, Čábelková, & Volicer, 2018), whereas Burgess et al. (1998), like us, report on healthy adults. Also, the proxies in our study are rather heterogeneous, varying in the different proximities to the participant, and are thus somewhat less reliable. In this case, it seems likely that individuals have a more accurate insight into their cognitive failures than do their proxies, who might not notice every occasion a mistake is being made, especially as this is a healthy sample with little cognitive complaints. Yet, it is also possible that individuals simply overestimate their own failures, judging themselves too negatively.

Floyd and Scogin (1997) indicated that effect sizes of subjective cognitive measures after training are usually much lower compared to those of objective tasks and that the correlation between the two is generally small. Indeed, in our exploratory results, we did not find a correlation between subjective and objective tasks. Additionally, in an earlier publication, we reported improvement on several objective tasks in all groups after the current training (Buitenweg et al., 2017). Although still fairly low, larger effect sizes were seen than on the subjective measures. A plausible explanation for this occurrence is that the objective tasks used in our study were more sensitive to retest effects compared to the subjective questionnaires.

Numerous intervention studies have included measures of depression or mood to gauge change after training, yet most fail to find improvement on these factors (Fairchild & Scogin, 2010; McDougall et al., 2010; Slegers et al., 2008; Wolinsky et al., 2006). Likewise, our cognitive flexibility training also did not affect participants’ depression levels. This could be due to the fact that depressive symptoms in our sample were low. There is some evidence that cognitive training could be of benefit to depressive symptoms in clinically depressed populations (Alvarez, Sotres, León, Estrella, & Sosa, 2008; Calkins et al., 2015), although these studies have only been done with younger adults. In older, clinically depressed populations, it is likely physical training might be more suitable to relieve symptoms (Blake, Mo, Malik, & Thomas, 2009), though to our knowledge, as of yet no effective cognitive training exists that indicates positive elevation of mood in older adults.

Regarding depression, we also observed a moderate correlation between depression scores and memory complaints, whereas a correlation between memory complaints and actual memory scores failed to appear, confirming earlier findings (Hertzog, Dixon, & Hultsch, 1990; Hertzog & Pearman, 2013; Hulur et al., 2015). One explanation is that feelings of mental distress or dysphoria affect individuals’ judgment, causing negative thoughts about themselves to overshadow impartial perception. For instance, on the subjective memory scale, we asked about memory complaints and worries. Especially individuals who are prone to negativity might have reacted differently to these questions than if we had posed them in terms of more positive terms, such as memory...
capacity. Another interpretation is that many older adults are not well capable of judging their own memory abilities. Crumley, Stetler, and Horhota (2014) argue that the ability to subjectively judge one’s own memory capacity does not start until one’s memory is actually declining, often in oldest-old adults, and that when assessing memory performance in a healthy sample of younger-old adults, the use of subjective measures should be discouraged. Indeed, in the current study, we included few oldest-old adults, and it’s possible that without many actual memory failures, most were not able to accurately evaluate their daily memory performance. In future studies of subjective memory and depression, therefore, it is essential to also include individuals from the highest end of the age spectrum.

Some caution is required in judging the overall reliability of the questionnaire results. For instance, subjects were mostly free to complete the questionnaires at a time point of their choice, thus giving way to more external influences. Although access to the list of questionnaires was only granted from a specific moment onwards, due to the online nature of measurement, subjects were permitted three days or more to fill them out, possibly allowing unrelated events or fluctuations of mood and fatigue to affect their reports. This is backed up by Krueger and Schkade (2008) who argue that subjective well-being is often no more reliable than a typical mood questionnaire, as respondents inadvertently use arbitrary and transitory information, such as mood, to evoke an impression of their general well-being. Therefore, despite using reliable measures, we have to remember that subjective self-report questionnaires show feelings experienced at a given moment in time.

Another limitation of the current study is that the level of autonomy we offered participants was low. Loos and Zonneveld (2016) argue that a certain amount of pressure from a game or intervention can contribute to the experience of working toward a goal, while too much pressure is thought to lower motivation (Ryan & Deci, 2000). In our game protocol, in order to ensure all participants followed the same order and number of different games and time spent on each domain, we pre-programmed each training session to include the games in a specific order. It is possible that by preparing a set range of games for our subjects, we potentially eliminated participants’ needed autonomy in all groups, thereby reducing overall internal motivation and enjoyment. Our previous comments on the training protocol in Hardy et al. (2015) apply here as well. One way to go about this in future studies is allowing for self pacing (Callahan, Kiker, & Cross, 2003) by granting participants more freedom in selecting the tasks and duration of engagement within each session, and relying more on their own capability for setting boundaries, possibly with use of reinforcement learning and feedback strategies (Green & Bavelier, 2008; Simon, Howard, & Howard, 2010). It is not obvious, however, what the ideal balance is between motivating freedom of choice, on the one hand, and limitations imposed by experimental rigor, on the other.

Finally, participants in this study were high functioning older adults, which becomes evident from the self-ratings on cognitive failures and executive dysfunction and the ceiling-level subjective independence scores. This may have made it more difficult to achieve further improvement on the constructs measured here. In future research, stronger effects might be found in individuals with subjective cognitive complaints or Mild Cognitive Impairment.
In addition to the aforementioned, it remains important that future research focuses on identifying those subgroups for which specific interventions might be most effective. We further advise that in future studies of cognitive training in older adults, attention is paid to offering individuals certain guidelines to foster intervention effects, for instance in using strategies in daily life (Wolinsky et al., 2006). Finally, it is important that before the start of a training session, essential pre-training, such as practice on the training games, is given to enhance both training results and internal motivation (Floyd & Scogin, 1997).

The current study was designed to answer the question whether a commercially available, state-of-the-art, adaptive cognitive flexibility training could lead to improved self-perceived cognitive and executive functioning, quality of life, depression, and anxiety. Though 139 subjects each completed 58 half-hour sessions of brain training during 12 weeks, we found no evidence of advantageous outcomes to support this claim, though many questions remain. We strongly encourage further research into cognitive training effects on subjective performance, in order to be able to close these gaps in our knowledge.

Acknowledgments

The authors would like to thank all participants and their relatives for participating in the study; all students for assisting in recruitment, testing, and coaching of participants, and Dezel media for making Brainymer available for our study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This project is part of the research program “Evidence-based adaptive brain training in seniors: Effects of brain structure and dopaminergic system on individual differences in trainability” funded by Brain, Cognition, Behaviour (HCG), a division of the Netherlands Organisation for Scientific Research (NWO), under grant number [056-12-010].

References


