Journal of Visualized Experiments

Working memory training for elderly – the importance of the control group's training regimen and initial intellectual functioning assessment.

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE60804R3
Full Title:	Working memory training for elderly – the importance of the control group's training regimen and initial intellectual functioning assessment.
Section/Category:	JoVE Behavior
Keywords:	Cognitive training, working memory, working memory capacity, working memory training, old adults, elderly population, predictors of cognitive training effectiveness, dual n-back
Corresponding Author:	Aneta Brzezicka POLAND
Corresponding Author's Institution:	
Corresponding Author E-Mail:	abrzezi2@swps.edu.pl
Order of Authors:	Olga Matysiak
	Wanda Zarzycka
	Aleksandra Bramorska
	Aneta Brzezicka
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Warsaw, Poland

TITLE:

Working Memory Training for Older Participants: A Control Group Training Regimen and Initial Intellectual Functioning Assessment

AUTHORS AND AFFILIATIONS:

Olga Matysiak[1], Wanda Zarzycka[1], Aleksandra Bramorska[1], Aneta Brzezicka[1,2]

[1]Department of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland

[2]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Email addresses of co-authors:

Olga Matysiak (omatysiak1@swps.edu.pl)

Wanda Zarzycka (wzarzycka@st.swps.edu.pl)

Aleksandra Bramorska (abramorska@st.swps.edu.pl)

Corresponding author:

Aneta Brzezicka (abrzezi2@swps.edu.pl)

KEYWORDS:

Cognitive training, working memory, working memory capacity, working memory training, old adults, elderly population, predictors of cognitive training effectiveness, dual n-back

SUMMARY:

A cognitive training intervention in elderly population together with the assessment of the pre training cognitive abilities is presented. We show two versions of training – experimental and active control – and demonstrate their effects on the array of cognitive tests.

ABSTRACT:

The efficacy of cognitive training interventions is recently highly debated. There is no consensus on what kind of training regimen is the most effective. Also, individual characteristics as predictors of training outcome are still being investigated. In this article, we show the attempt to address this issue by examining not only the impact of working memory (WM) training on cognitive training effectiveness in older adults but also the influence of the initial WM capacity (WMC) on the training's outcome. We describe in detail how to perform 5 weeks of an adaptive dual n-back training with an active control group (memory quiz). We are focusing here on technical aspects of the training as well as on the initial assessment of participants' WMC. The evaluation of pre and post training performance of other cognitive dimensions was based on the results of tests of memory updating, inhibition, attention shifting, short-term memory (STM) and reasoning. We have found that the initial level of WMC predicts the efficiency of the n-back training intervention. We have also noticed the post training improvement in almost all aspects of cognitive functioning we measured, but those effects were mostly intervention independent.

INTRODUCTION:

In many cognitive trainings studies, the dual n-back task is used as a method of working memory (WM) training. WM is a common target of cognitive interventions because of its importance for other, higher level intellectual functions[1]. However, the effectiveness of such training and its potential for creating a more general improvement in cognition, has been highly debated (for meta-analysis, see[2–7,14] and for reviews, see[4,8–13]). While some researchers claim that ''… there was no convincing evidence of the generalization of working memory training to other skills''[4], others present meta-analytical data, which show highly significant effects of WM training[2,3,5,6,11]. The separate problem is the effectiveness of WM in the elderly population. Several WM training studies reported greater benefits in younger adults compared to older adults[15–20], whereas others show that similar effects can be observed in both age groups[21–25].

Various elements are believed to forecast the benefits of memory training[26]. Some of those factors appears to be potential moderators of WM training effectiveness[21]. Mental capacity, being described as the baseline cognitive capacity or general cognitive resource, seems to be one of the strongest choices for this position. In order to assess the role of the initial intellectual level, we put a special emphasis (the method described here) on the measurement of the cognitive capacity before applying a training regimen. It was dictated by the data showing that participants, who are characterized by higher cognitive capacity at the beginning of the training, achieving substantially better training outcomes compared to those with lower levels of initial cognitive functionning[27]. A similar phenomenon is observed in educational research, where it is referred to as the Matthew effect[28], an observation that people with initially better skill improve even more when compared to those with preliminary lower level of ability in question.

It is thought-provoking, though, that not so many reports have been published on this topic[21,29]. Moreover, even substantial individual differences, especially when it comes to the elderly population, are often left unattended during data analysis and interpretation[30]. In the present study, we examine the impact of the initial level of working memory capacity on WM training success in the group of healthy older adults. In order to maintain every element of the training regimens as similar as possible between experimental and control groups, we employed an active control group design. Therefore, the training content (WM versus semantic memory) remained the one crucial factor determining the expected difference in the training results. Both groups performed computerized, home-based trainings. Members of the experimental group were assigned to an adaptive dual n-back training program and an active control group trained with a task based on a semantic memory quiz. New in the approach here is the emphasis on the initial evaluation of the participants' cognitive level by assessing their working memory capacity (WMC). Additionally, the method of assessing the initial WMC level we present in this article has proven to be an effective tool in distinguishing between people who will and will not be successful during subsequent working memory training. We have previously described and published results from this study[44]. Therefore, in this article we are focusing on a detailed description of the protocol we used.

PROTOCOL:

The SWPS University of Social Sciences and Humanities Ethics Committee assessed the protocol described here. A written informed consent in accordance with the Declaration of Helsinki was obtained from every participant.

1. Participants recruitment

1.1. Recruit at least 36 volunteers for each training group. This number was evidenced to be enough to observe between-groups effects by authors' previous research and also in the literature on the subject. Typical effect sizes in working memory trainings studies range between d=0.6 to 0.8 depending on the training type or targeted group. Based on these values and aiming at a decent statistical power of 0.8 with alpha at 0.05, a minimum sample size in this type of research (calculated according to a formula proposed by Soper[45]) is 36 (preferably more).

1.1.1. Use the following inclusion criteria: over 55 years old, with no history of neurological or psychiatric disorders, with preserved motor skills of the upper limbs, without blindness or hearing loss, who are not presently involved in any other cognitive (especially memory) training.

1.2. Recruit participants using various methods: online advertisements posted on social media profiles, research and work platforms or discussion groups, as well as in-person announcements at Universities of a Third Age or during events involving older audience such as picnics for seniors (also with usage of posters and leaflets) to get certainty that the people who are recruited are not only internet users.

1.3. Remember to adequately describe in the advertisement what type of participants you are looking for.

2. The evaluation of the Ethics Committee

2.1. Before starting the study, obtain the evaluation form your local Ethics Committee, including permission to: a) Interactions consisting of active intervention in human behavior, aimed at changing this behavior, without directly interfering with the brain, e.g. cognitive training, psychotherapy, etc. (this also applies to intervention intended to benefit the respondent, e.g. improving his memory), b) collect and process participants' personal data, in particular data that allows to identify the subjects.

3. Initial screening

3.1. Begin with a short interview during informing the participant in detail about the project's aim, the possibility of withdrawal and personal data protection.

3.2. Be sure that participant does not take drugs or has never suffered from any disease that may affect central nervous system functioning. Likewise, control intake of medications not related to neurological diseases that affect cognitive functioning. If the screening revealed any unwanted information, exclude the volunteer from the study.

3.3. After successful screening, present the written informed consent form to the participant and ask them to read it. The written informed consent should involve information as follows: a) the legal bases for data gathering and processing specific for a given country, b) information about rights of the data owner (e.g., accessing the personal data, possibility to supplement incomplete personal data, deleting data or processing restrictions).

3.3.1. Ask the participant to sign the informed consent.

3.4. Carry out the Mini Mental State Examination (MMSE)[32] to ensure that participant shows no signs of mild cognitive impairment – at least 27 points are needed to enter next stages of a study.

3.4.1. Read the MMSE's Introductory Script to the participant and then ask questions in accordance with the examination script.

3.4.2. Begin with an assessment of orientation to place and time by asking series of questions: What is today's full date? What day of week is today? Where are we (what city, name the building, which floor)?

3.4.3. Follow with memory test: ask participant to memorize three objects that was read at loud by you; go through the series of seven tasks assessing attention, concentration, and calculation, and at the end ask participant to recall the three objects previously learned.

3.4.4. Finally, test naming, repetition, and understanding according to examination script.

3.4.5. Score responses as follows: 0 = incorrect or lack of an answer, 1 = correct answer.

3.4.6. Do not take longer than 10 min for the administration of the MMSE test.

3.4.7. If a person does not reach the required threshold (27 points), inform them of the result. If there is any suspicion of a clinically lowered level of cognitive functioning, refer such a person to a specialist unit (e.g., a certified psychologist at a neurological center).

3.5. Store documentation in a manner that complies with the law and/or The General Data Protection Regulation of the local country.

4. Training group assignment

4.1. Randomly assign participants to an experimental or control group (**Figure 1**). To ensure the randomness of the process, generate a list of 50 codenames (**Figure 2**) assigned to ones and twos (training groups) and connect each participant in recruiting order with those codes (saved in separate file). From now on, replace participant data with codenames.

--- Figure 1 around here ---

4.2. Make sure that the group assignment is not biased in terms of age, gender or the level of education. Pre-assign age, gender and education group to the list of code-names for each training group (1 and 2), as presented in **Figure 2**. Fit each volunteer in the table based on their characteristics.

--- Figure 2 around here ---

5. Initial assessment of cognitive functioning

5.1. Put an emphasis on giving very clear and detailed instruction to the participants about how to go through each task at the beginning of each procedure. Run the "training block" before each task (that is identical to the training task) and observe the participant if their answers indicate that they understand the instructions.

NOTE: The inclusion of such blocks is described at the beginning of description of each task below.

5.2. After presenting the instruction and practice part of each task and before starting the main part of the procedure, once again ask if the participant understands the procedure's requirements.

5.3. Be sure that each participant will accomplish the full set of the following tasks.

5.4. The operation span task (OSPAN)

5.4.1. Run a training block to estimate the individual time needed for each participant to calculate a simple mathematical equation (adding, subtraction, dividing and multiplication of single digits - not higher numbers).

5.4.2. In the middle of the white screen display the equation to the participant. Instruct the participant to think of the result and then press an arrow leading to the next screen, where the outcome of equation is presented. Let the participant give an answer by pressing either the **True** or **False** button.

5.4.3. Count the time needed for resolving the equation. Use mean time needed in a final block as a time of displaying the equation in the main part of the task. Have a fixed time limit imposed for estimating the correctness of the equation result: 5 s.

5.4.4. In the next training block display letters on the screen - one by one, for 500 ms each, and instruct the participant to remember them. After a full set (3 to 9 letters), present the matrix of 12 letters to participant and ask to mark memorized letters in a correct order. Do not give a time limit for the answer. Record the correctness.

5.4.5. Run a main part of the task. In the final block, mix two abovementioned training blocks: after each part with equation (remember about time limits!) present one letter to remember. Display 2 to 5 pairs of 'equation + letter' and then, after presenting the whole sequence of 'equation + letter' pairs, show the matrix of letters for participant to mark memorized letters in a correct order. Record the correctness of mathematical and memory part.

NOTE: With this test the operational span of Working Memory (processing the one kind of information while remembering the other) is assessed.

5.5. The Sternberg task

5.5.1. Present a random sequence of digits (2 to 5 in a one set) in white font on black screen, for 500 ms each, with the interval ranging from 2500 to 3000 ms.

5.5.2. Display a fixation cross for 2500 ms.

5.5.3. At the end of presented sequence, display a target digit in a yellow font for 500 ms.

5.5.4. Let the participant decide if the yellow digit appeared among the presented earlier set of white digits by pressing **Yes/No** buttons. If the participant does not give an answer within 3000 ms, go to the screen with fixation point and start the next trial. Count this attempt as a wrong answer.

5.5.5. Repeat steps 5.5.1 to 5.5.4 120 times (trials) with 50% of probes containing target digit in the sequence and 50% do not (randomly).

5.5.6. Record the correctness and the reaction time for each trial.

NOTE: This task tests the speed of searching the information in memory. The increase of reaction time accompanies the enlargement of the set, which is explained as the process of a serial search of the memory content.

5.6. The running memory span task

5.6.1. On the screen, present information about the number of letters to remember (3, 4, 5 or 6 letters depending on the difficulty level of a block) and ask the participant to go to the next screen by pressing a key.

5.6.2. Present a sequence of letters, one by one, in black font in the center of a white screen, for 0.25 s each.

5.6.3. Ask the participant to reproduce a fixed number of the last letters from the sequence: fixed no: 4; sequence; K U J D S T W A; letters to memorize: S T W A.

5.6.4. To receive the participant's answer, display on the screen a matrix of 9 letters (3x3) and ask the participant to mark appropriate letters (in order the letters appeared) with the mouse. Do not give a time limit for the answer.

5.6.5. Record correctness (mind the sequence errors).

NOTE: This test measures the capacity of the working memory with the use of the additional distraction in the form of the inability to predict which letters from the list would be the portion to remember.

5.7. Test Go-No Go

5.7.1. On the white screen display trials composed of: a) 250 ms – a fixation point (white cross), b) 1250 ms - the stimuli (a letter), c) 2000 ms – a fixed inter-stimulus interval.

5.7.2. Have the participant react by pressing a key as quickly as possible, when a target stimulus – letter X – appears on the screen.

5.7.3. Record the reaction time and correctness of answers.

NOTE: The test measures the efficiency of inhibition in two conditions: in easier condition the letter X is presented in the 50/50 proportion to other letters and in the more difficult condition the target stimulus is displayed in 70/30 proportion to other letters.

5.8. The switching task

5.8.1. Divide the screen into two parts using a horizontal line. Present red squares or rectangles composed of smaller squares or rectangles above or below this line.

5.8.2. Apply two different rules for the participant to react, depending on where figures will appear - "pay attention to the small figures" (local) for an upper part of the screen and "pay attention to the whole figure composed from smaller figures" (global) for a lower part of the screen. Have the participant react according to the part of the screen where the stimuli appeared.

5.8.3. Add a cue indicating to which dimension (global or local) the participants should respond. Cues related to the local (global) dimension should consists of a small (big) red square, presented at one side of the target stimulus, and a small (big) red rectangle, displayed at the other side of the target stimulus. Make sure that the cue remains on the screen until a response is given.

5.8.4. Display the figures above or below the middle line in random order.

5.8.5. Have the participant react according to the previously presented rules: answering “rectangle” by using the left button and answering “square” by pressing the right one.

5.8.6. Record reaction times and correctness of the answers.

NOTE: Time for the response have to be fixed at 3500 ms. The time interval between the cue and the target stimulus should be 500 ms. The interval between the response and the presentation of the cue should be 1000 ms. Each figure and each cue should be presented for all the time needed for the participant to react by pressing one of the keys. The Switching Task measures the cognitive fluency, as it requires quick attention switching between the respective elements.

5.9. The linear syllogism task

5.9.1. Display on the screen a set of three ‘premises’ that together constitute logical chain of relations: pairs of letters with an information about a relation between them: A > B, B > C and C > D. Each premise should be visible on the screen for 1500 ms and interval between them should last 3000, 3500 or 4000 ms (randomly). An integrated mental model representation[32] of such set of pairs will always be in the linear order “A > B > C > D”.

5.9.2. Include three pairs of possible relations between premises in separate trials: 1) A > B, B > C, C > D (adjacent pairs, exactly the same as those that had been seen in the learning phase), 2) A > C, B > D (two-step relations, not seen before and requiring integration of information), 3) A > D (end point relation, not seen before and requiring integration of information).

5.9.3. Construct the task so that it contains two conditions: an easy condition, where the premises should be displayed one after the other in the order in which they form a logical string (e.g. string: Q>W>E>R>T, premises order: Q>W, W>E, E>R, R>T); a difficult condition, the order of the premises should be altered (e.g. premises order: W>E, Q>W, R>T, E>R).

5.9.4 Test the participant immediately after the presentation of premises by displaying statements (for 1500 ms each) that participant needs to assess as true (answer: right button) or false (answer: left button), as fast as possible. Set time limit for the answer - 6000 ms, and after each answer given wait an additional 1000 ms before displaying the next question. Each statement should consist of only one pair of letters and relation between them (‘<’ or ‘>’) in either a correct (e.g., “W > E ?”) or false setup (e.g., “E > W ?”).

5.9.4. Randomize the arrangement of the letters in order to minimize possible interference induced by implied alphabetical order of letters.

5.9.5. Use capital letters as stimuli instead of whole sentences in order to avoid linguistic connotations, and the symbol “>”to indicate the relation between elements.

5.9.6. Gather the data about the accuracy and reaction time of the response for each question.

NOTE: Questions about adjacent pairs are used to estimate memory, and questions about premises presented in a mixed order, and those asking about the relationship between the far elements of logical sequences are asked to measure the ability to information integration.

6. Training protocols

6.1. For both experimental (n-back) and control (Quiz) trainings provide participants with the access to the Internet platform (logins and passwords) – which allowed them to enter the site every 24 hours, to avoid situations in which the participant trains more than once a day

6.2. Make sure that participant understands the task as well as the training regimen.

6.3. Instruct participant to train in similar conditions every time, in calm and quiet place with possibly low level of external distractors.

6.4. Experimental training: working memory paradigm

NOTE: An adaptive dual n-back task served as a working memory training program. This task was introduced by Jaeggi et al.[33] and simultaneously recruits auditory and visual attention, maintenance, and updating processes.

6.4.1. Instruct participant on the task at level N=2 (see **Figure 1B**).

6.4.2. Use alphabet letters as auditory stimuli and green squares, presented in one of nine locations in a 3x3 matrix, as visual stimuli.

6.4.3. Present a single item for 500 ms followed by 2500 ms interval, during which the participants are supposed to respond. The current stimuli can match the target visual (response with left hand) or auditory stimulus (response with right hand) or both (response with both hands simultaneously).

6.4.4. Single session of the n-back training

6.4.4.1. Set level of N to 2, in the first block of the task. After each block, evaluate the correctness of the answers and, on this basis, adjust the level N in the next block. If the accuracy surpassed 85%, the difficulty level should be increased (by 1 point), if the accuracy drops below 60%, lower the difficulty level. In other cases the N remains unchanged.

6.4.4.2. For the first block, fix the n-back task level at N=2. Later, determine the N level for the current block based on the correctness of answers in the previous block. If the accuracy exceeds 85%, increase difficulty level. If the accuracy is below 60%, the level of difficulty should be lowered. In other cases, the N level should remain unchanged.

6.4.4.3. Set a single dual n-back session for 15 rounds (15 blocks of tasks), each with 20 + N trials and the whole training set for 25 sessions.

6.4.4.4. Record reaction times (RTs) and accuracy (ACC) measures for each trial.

6.5. Control training: episodic memory paradigm

6.5.1. Collect material from the Internet in order to construct the Quiz Task which engages semantic memory (e.g., what is the capital city of Hungary?).

6.5.2. Present 15 questions in each training session of Quiz Task (starting from the second session 5 questions come from the previous session and 10 new) with no time limit for reading it. Instruct participant that after selecting the 'answer' button they have to choose one of the four given possibilities within 40 seconds. Provide the feedback for correctness of the answers.

6.5.3. Set the whole training for 25 sessions.

7. Training supervision

7.1. During the course of the training verify the training progress of each participant. Assign an experimenter who is responsible for checking (online) the progress of the training, to each participant.

7.2. If a break between sessions is longer than two days, have the experimenter contact the participant via text message and encourage him or her to resume the training.

8. Post-training assessment of cognitive functions

8.1. Proceed with post-training session in the exact way as the pre-training meeting.

8.2. Compensate each participant, who completed the whole protocol for the time dedicated to the research, with 150 PLN (~$40).

REPRESENTATIVE RESULTS:

Training-related effects

85 subjects participated in the study (29 were male) and they were on average 66.7 years old. Due to technical problems, data from one participant in the n-back training group was not recorded. Finally, the data from 43 participants in n-back training group and 42 in Quiz training group were analyzed. Multivariate analysis of variance (MANOVA) with repeated measures was used to analyze training specific effects for both training groups over time (pre-, post-training). The results of each cognitive test were dependent variables (**Table 1**), and training group and

measurement points (pre- versus post-training) were independent variables. These results are presented in **Table 2**.

The results of the analysis indicated a statistically significant post-training improvement in the syllogisms task: ($F(1,83)=31,22$, $p<0.001$, $\eta p2=0.27$) and attention switching task: ($F(1,83)= 5.79$, $p=0.02$, $\eta p2=0.07$). A significant training group effect was observed for memory SPAN task ($F(1,83)=7.72$, $p=0.01$, $\eta p2=0.09$) and OSPAN task ($F(1,83)=13.01$, $p=0.01$, $\eta p2=0.14$). None of the interaction effects (time x training group) has proven to be statistically significant. However, we found significant within group effects for some analysis. In the OSPAN task, the n-back training group improved their results in second session), while for the quiz group, both performances were similar. This effect needs to be interpreted in reference to the fact that the quiz and the n-back group differed in the first measurement. Thus, results of the n-back group where initial OSPAN performance was higher improved, while the control group did not. The performance in Sternberg's and a go/no-go task was not related to a training group assignment or the time of measurement.

Overall, the results show that participants' cognitive performance was improved in the post-training execution of attention and higher cognitive functions (reasoning) engaging tests, regardless of the group affiliation.

--- Table 1 and 2 around here ---

WMC as a predictor of WM training effectiveness
In a subsequent analysis, performed on the n-back training group only, we used a more refined method - multilevel modeling (MLM) - to observe the learning process during the experimental training. The hierarchical structure of the data was accommodate to the model: at level 1 - repeated measurements, nested within participants (level 2)[34]. The MLM dataset consisted of 1,050 observations gathered from 42 participants from experimental group within each of 25 training sessions. The model provided for both fixed and random effects: the regression intercept and slope for the average person, and between-subject variability around the average. In the Model 1, the change in the number of points scored in the n-back task over time (number of the training sessions) was modeled. The time variable was centered at 1st day of the intervention. Compared to Model 1, Model 2 added on predicting and moderating effects of a baseline OSPAN score (between-subjects predictor - level 2) on within-subject variability (level 1). Those predictors were tested independently to avoid multicollinearity. In all models, a linear and quadratic effects for the slope were tested, however the quadratic one was subsequently removed because its fixed effects and variance components were not significant. The restricted maximum likelihood served as the estimator. –2 Restricted log likelihood ratio (-2LL) and the Akaike Information Criterion (AIC) were used to assess the goodness of fit for all models. Given the common proximal autocorrelation in the daily data[35] we decided to base on a first-order autoregressive [AR(1)] covariance structure.

MLM results showed that OSPAN scores from the pre-training measurement were a significant predictor of the first n-back outcome from the 1 session. Baseline OSPAN level turned out to be a moderator of the whole training course (**Table 2**). When compared, groups of participants with high or low OSPAN points had similar N-level at the first training session: approx. 2.00 units on a 1+∞ scale (low OSPAN = 1.93; high OSPAN = 2.31). A significant difference manifested in the post-training measurement, when the participants with low initial OSPAN results achieved a .01 unit increase in n-back task, whereas those with the high initial OSPAN scores recorded a .04 point improvement. The observed result clearly indicates the existence of a positive association between the initial OSPAN level and training effectiveness. The n-back scores in the 1st session and a learning curve of a training are higher and stepper for the participants with initially higher OSPAN result ($p < .001$).

--- Table 3 around here ---

FIGURE AND TABLE LEGENDS:

Table 1. Descriptive statistics for the cognitive tasks' results.

Table 2. Outcome measures: main and interaction effects from MANOVA with training type (n-back vs. Quiz) and time (pre vs post training) as factors.

Table 3. Multilevel analysis of the training data (n-back task performance as a dependent variable). Models with training sessions (time) only (MODEL 1) and training sessions plus initial working memory capacity level as a predictor and a moderator, respectively (MODEL 2).

Figure 1. Study design with examples of a training tasks. Participants underwent two measurement sessions, before and after a 5 week training protocol.

Figure 2. The example of suggested coding form for group assignment.

DISCUSSION:

In the study presented here, we have investigated whether older adults could benefit from working memory training and if it is connected to the initial level of their basic cognition. We used an n-back task as an experimental intervention and working memory capacity (measured with the OSPAN task) was the method to probe participants' initial level of intellectual functioning. We had two critical steps in the protocol. The first and most important was the assessment of the initial WM level. The second was the careful matching of the two training regimens in every possible way but the "cognitive content" (i.e., working memory versus semantic memory). By introducing the assessment of the participants' cognitive level at the beginning of the study, we were able to show how important it is to have a good estimate of it at the start of the intervention. It was the most important predictor of the cognitive training's effectiveness. We suspect that in most intervention studies, researchers assess, in one way or another, the initial cognitive level of participants. To obtain such information it is possible to use the results from the first measurement of a trained cognitive task as a predictor of the cognitive training effectiveness. As expected, the N level of n-back task fluctuated substantially through training sessions. What was even more interesting, individuals with higher maximal N achieved in the first training session improved faster than the rest of the group in the following sessions. That implies that the variability in performance between participants, noticed at the beginning of the study, only increased with time and training. To explore this effect deeper we conducted further analysis. The results showed the preliminary score in OSPAN task (WMC) to be a strong predictor of the improvement gained during the training course (in the dual n-back task). Participants characterized by higher initial WMC performed better in the training from the very first day and had stepper learning curve in comparison to seniors with WMC below the average of the sample. We are not the first to report such effect. Foster et al. (2017) described similar results[29]. They proved the existence of the correlation between the initial WM level and the performance of memory span training. This result is consistent not only with the ones here, but also with research on the so-called Matthew effect in WM training interventions, in which participants with initially higher skills profit more from training and score better in both: trained and untrained, tasks[21,36–39]. All this strengthens the conclusion that someone's ability to gain from WM training depends heavily on the initial intellectual level.

Regarding the regimens similarity, we applied the Mill's method of one difference[40:] : when someone observes one situation that leads to a given effect, and another that does not result in the same way, and the only difference between these situations is a presence of a specific factor only in the first situation (here, the difference in the cognitive layer), there is the solid foundation to assume that it is the factor in question that caused the observed effect. We tried to match the training regimens in terms of motivation, superficial similarities (same amount of training sessions, similar feedback, etc.). It is worth noting that the first idea was to use the same task (n-back) but in its easiest form, where the N level is fixed to 1. It quickly becomes obvious that it was a wrong path as the participants (in pilot studies) not only reported weariness but also were dropping off the control condition at a much higher rate than from the experimental (with adaptive level of difficulty). This resulted in a different approach. After several pretests we decided for a "different function" approach (WM versus semantic memory) instead of having the same function in both conditions just with different intensity (fixed level of WM versus adaptive level of WM). One potential problem with such approach is that we can create a control condition, which is more attractive than the experimental condition. And, if motivation to engage is a crucial factor in cognitive trainings, we can have null results because of that decision.

It is worth noticing that there is a substantial change in a way we look now at results from cognitive intervention studies. For example, Reddick et al. suggest that positive effects observed in WM training groups when compared to control groups are due to decrease present in control group and not improvement of performance in experimental groups[41]. When we think about elderly population, even such output – maintaining of the initial cognitive level – could be a desirable outcome. But, surprisingly, in the study we did not observe a post-training reduction of performance in the control group, except for the go/no-go task. It might be, again, interpreted as evidence that even a simple memory quiz, if it is attractive and encourages participants to engage in some cognitive activity, could produce beneficial effects. What is also important, all of the participants (regardless of group assignment) volunteered for the study and some correlational studies have shown that voluntary work might be a protective factor against cognitive aging[42,43]. One of limitations of the study is that we do not have the representative population of older people. Instead, the elderly taking part in the study were probably more motivated and more proactive than seniors who, for example, do not leave their homes. However, the level of education and economic status (indirectly controlled - as an occupational activity that generate income) were measured in the study and the analysis showed that these were not factors affecting training progress. It can be also argued that improvement observed in both interventions is the result of mere test-retest effects. Due to the fact that there was no passive control group included in the study, this matter cannot be settled down in this study. It is therefore advisable to include another group in the subsequent tests - passive control. The most important message from the study is that the findings suggest that post-training gains are within reach of older adults, especially those characterized by a good overall cognitive functioning. What we wanted to delineate in this article was the way we were introducing and maintaining the participants in a training regimen. The most important thing in this study was to keep all features of the intervention exactly the same between the two groups apart from one thing – the cognitive function involved undergoing practice. As we did not observe substantial differences between the effectiveness of the training protocols, but the improvement was visible in both groups, it seems valid to conclude that any cognitive engagement can be beneficial for elderly people. As the main result refers to the initial level of cognitive functioning, we strongly recommend including initial measures of the trained function and verifying it as a possible predictor (or at least co-factor) of training effectiveness.

ACKNOWLEDGMENTS:

Described results are obtained from the project supported by the National Science Centre in Poland under grant no. 2014/13/B/HS6/03155.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1. Baddeley, A. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* (2003).
2. Au, J. et al. Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin and Review*. **22**, (2) 366–377 (2015).
3. Karbach, J., Verhaeghen, P. Making working memory work: A meta-analysis of executive control and working memory training in younger and older adults. *Psychological Science.* **25**, (11)2027–2037 (2014).
4. Melby-Lervåg, M., Hulme, C. Is working memory training effective? A meta-analytic review. *Developmental Psychology* (2013).
5. Melby-Lervåg, M., Redick, T. S., Hulme, C. Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science* (2016).
6. Schwaighofer, M., Fischer, F., Bühner, M. Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. *Educational Psychologist* (2015).
7. Soveri, A., Antfolk, J., Karlsson, L., Salo, B., Laine, M. Working memory training revisited:

A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review* (2017).

8. Au, J., Buschkuehl, M., Duncan, G. J., Jaeggi, S. M. There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychonomic Bulletin and Review* (2016).
9. Dougherty, M. R., Hamovitz, T., Tidwell, J. W. Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin and Review* (2016).
10. Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., Schmiedek, F. A Theoretical Framework for the Study of Adult Cognitive Plasticity. *Psychological Bulletin* (2010).
11. Shinaver, C. S., Entwistle, P. C., Söderqvist, S. Cogmed WM Training: Reviewing the Reviews. *Applied Neuropsychology: Child* (2014).
12. Shipstead, Z., Redick, T. S., Engle, R. W. Does working memory training generalize? *Psychologica Belgica* (2010).
13. Shipstead, Z., Redick, T. S., Engle, R. W. Is working memory training effective? *Psychological Bulletin*. **138**, (4)628–654 (2012).
14. Weicker, J., Villringer, A., Thöne-Otto, A. Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology* (2016).
15. Brehmer, Y., Westerberg, H., Bäckman, L. Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience* (2012).
16. Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., Nyberg, L. Transfer of learning after updating training mediated by the striatum. *Science* (2008).
17. Dorbath, L., Hasselhorn, M., Titz, C. Aging and executive functioning: A training study on focus-switching. *Frontiers in Psychology* (2011).
18. Heinzel, S. et al. Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology, and Cognition* (2014).
19. Schmiedek, F., Lövdén, M., Lindenberger, U. Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience* (2010).
20. Zinke, K. et al. Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology* **50**, (1)304–315 (2014).
21. Bürki, C. N., Ludwig, C., Chicherio, C., de Ribaupierre, A. Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. *Psychological Research* (2014).
22. Li, S. C. et al. Working Memory Plasticity in Old Age: Practice Gain, Transfer, and Maintenance. *Psychology and Aging* (2008).
23. Richmond, L. L., Morrison, A. B., Chein, J. M., Olson, I. R. Working memory training and transfer in older adults. *Psychology and Aging* **26**, (4)813–822 (2011).
24. von Bastian, C. C., Oberauer, K. Effects and mechanisms of working memory training: a review. *Psychological Research* **78**, (6)803–820 (2014).
25. Zając-Lamparska, L., Trempała, J. Effects of working memory and attentional control training and their transfer onto fluid intelligence in early and late adulthood. *Health Psychology Report*. **4**, (1)41–53 (2016).

26. Verhaeghen, P., Marcoen, A. On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging* (1996).
27. Borella, E., Carbone, E., Pastore, M., De Beni, R., Carretti, B. Working memory training for healthy older adults: The role of individual characteristics in explaining short- and long-term gains. *Frontiers in Human Neuroscience* (2017).
28. Bakermans-Kranenburg, M. J., Van Uzendoorn, M. H., Bradley, R. H. Those who have, receive: The matthew effect in early childhood intervention in the home environment. *Review of Educational Research* (2005).
29. Foster, J. L. et al. Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning Memory and Cognition*. **43**, (11)1677–1689 (2017).
30. Kliegel, M., Bürki, C. Memory training interventions require a tailor-made approach: Commentary on McDaniel and Bugg. *Journal of Applied Research in Memory and Cognition* (2012).
31. Sternberg, S. Memory-scanning: mental processes revealed by reaction-time experiments. *American Scientist* (1969).
32. Folstein, M.F., Folstein, S.E., McHugh, P.R. Mini-mental status. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. **12** (3): 189–98 (1975).
33. Jaeggi, S. M., Buschkuehl, M., Jonides, J., Perrig, W. J. Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America* (2008).
34. Bolger, N., Laurenceau, J. P. Methodology in the Social Sciences. (2013).
35. Kwok, O. M. et al. Analyzing Longitudinal Data With Multilevel Models: An Example With Individuals Living With Lower Extremity Intra-Articular Fractures. *Rehabilitation Psychology* (2008).
36. Bissig, D., Lustig, C. Who benefits from memory training? *Psychological Science* (2007).
37. Borella, E., Carretti, B., Riboldi, F., De Beni, R. Working Memory Training in Older Adults: Evidence of Transfer and Maintenance Effects. *Psychology and Aging* (2010).
38. Kraemer, K. R., Enam, T., McDonough, I. M. Cognitive reserve moderates older adults' memory errors in an autobiographical reality monitoring task. *Psychology and Neuroscience* (2019).
39. López-Higes, R. et al. Efficacy of cognitive training in older adults with and without subjective cognitive decline is associated with inhibition efficiency and working memory span, not with cognitive reserve. *Frontiers in Aging Neuroscience* (2018).
40. Mill, J. S. A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and the methods of scientific investigation, Vol. 1 (1875).
41. Redick, T. S. Working memory training and interpreting interactions in intelligence interventions. *Intelligence* (2015).
42. Guiney, H., Machado, L. Volunteering in the Community: Potential Benefits for Cognitive Aging. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* (2018).
43. Proulx, C. M., Curl, A. L., Ermer, A. E. Longitudinal Associations between Formal Volunteering and Cognitive Functioning. *Journals of Gerontology - Series B Psychological*

Sciences and Social Sciences (2018).

44. Matysiak, O., Kroemeke, A., Brzezicka, A. Working Memory Capacity as a Predictor of Cognitive Training Efficacy in the Elderly Population. *Frontiers in Aging Neuroscience*. **11**, 126 (2019).

45. Soper, D.S. A-priori Sample Size Calculator for Student t-Tests [Software]. Available from http://www.danielsoper.com/statcalc (2013).

Figure1

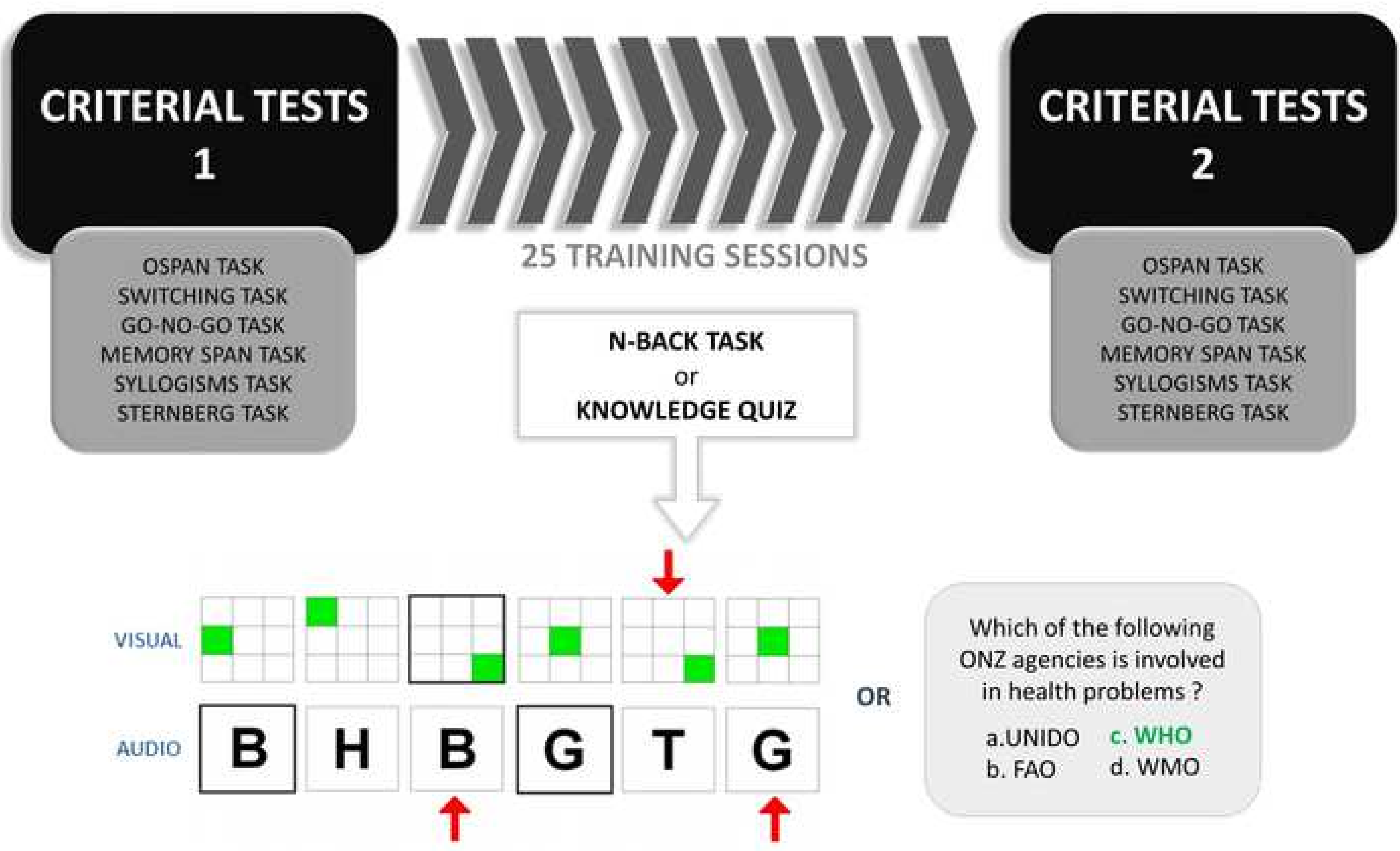
CRITERIAL TESTS 1
OSPAN TASK
SWITCHING TASK
GO-NO-GO TASK
MEMORY SPAN TASK
SYLLOGISMS TASK
STERNBERG TASK
25 TRAINING SESSIONS
CRITERIAL TESTS 2
OSPAN TASK
SWITCHING TASK
GO-NO-GO TASK
MEMORY SPAN TASK
SYLLOGISMS TASK
STERNBERG TASK
N-BACK TASK
or
KNOWLEDGE QUIZ
VISUAL
AUDIO
B
H
B
G
T
G
OR
Which of the following ONZ agencies is involved in health problems ?
a.UNIDO
b. FAO
c. WHO
d. WMO

Figure2

code	training	gender	age	education
name1	1	1	1	1
name2	1	1	1	2
name3	1	1	1	3
name4	1	1	2	1
name5	1	1	2	2
name6	1	1	2	3
name7	1	1	3	1
name8	1	1	3	2
name9	1	1	3	3
name10	1	2	1	1
name11	1	2	1	2
name12	1	2	1	3
name13	1	2	2	1
name14	1	2	2	2
name15	1	2	2	3
name16	1	2	3	1
name17	1	2	3	2
name18	1	2	3	3
name19	2	1	1	1
name20	2	1	1	2
name21	2	1	1	3
name22	2	1	2	1
name23	2	1	2	2
name24	2	1	2	3
name25	2	1	3	1
name26	2	1	3	2
name27	2	1	3	3
name28	2	2	1	1
name29	2	2	1	2
name30	2	2	1	3
name31	2	2	2	1
name32	2	2	2	2
name33	2	2	2	3
name34	2	2	3	1
name35	2	2	3	2
name36	2	2	3	3
etc.	...	...	...	...

Table 1. Descriptive statistics of the cognitive tasks results.

		N-back training				Quiz training		
	session	N	Mean	Std. Err	Std. Dev.	N	Mean	Std. Err
OSPAN task	1	42	15.31	1.64	10.62	40	9.07	1.77
	2	43	20.74	2.48	16.3	40	10	1.81
Syllogisms task	1	43	0.59	0.03	0.2	42	0.58	0.03
	2	43	0.67	0.03	0.21	42	0.69	0.03
Memory SPAN task	1	42	0.37	0.03	0.16	42	0.2	0.02
	2	41	0.4	0.03	0.18	42	0.22	0.03
Go/no-go task	1	42	0.14	0.05	0.33	42	0.16	0.03
	2	42	0.17	0.03	0.18	42	0.04	0.05
Sternberg's task	1	43	0.93	0.02	0.11	42	0.9	0.02
	2	43	0.94	0.01	0.05	42	0.93	0.01
Attention switching task	1	42	0.49	0.04	0.28	41	0.52	0.05
	2	42	0.41	0.04	0.23	42	0.46	0.04

Std. Dev.
11.22
11.45
0.21
0.19
0.16
0.18
0.01
0.12
0.15
0.07
0.3
0.25

Table 2. Statistical evaluation of the change in outcome measures (RM-MANOVA with training group (N-back and measurement points (pre- versus post-training) as independent variables and each test's outcome as dep

	Pre- to post-training effect			Training group effect			(ti	
	F (1,83)	ηp2	p	F (1,83)	ηp2	p	F (1,83)	ηp2
OSPAN Task	3.67	0.04	0.06	13.01*	0.14	0.01	1.49	0.19
Syllogisms Task	31,22*	0.27	0	0.01	0	0.95	0.35	0.01
Memory SPAN Task	3.13	0.04	0.08	7,72*	0.09	0.01	0.04	< .001
Sternberg's Task	3.56	0.04	0.06	0.78	0.01	0.38	0.62	0.01
Attention Switching Task	5,79*	0.07	0.02	0.75	0.01	0.39	0.02	< .001
Go/no-go Task	0.01	< .001	0.93	0.21	0.01	0.65	2.82	0.03

* statistically significant effect (p < .05)
T1 vs. T2 - difference in means between 1st and 2nd session;
N-back vs. Quiz - difference in means between training groups;

vs. Quiz)
endent variables).

Interaction effect (me x training group)	
p	significant within-group effects:
0.22	Nback (T1 vs.T2): 5,00*
0.56	
0.85	T1 (N-back vs. Quiz): 0,09* T2 (N-back vs. Quiz): 0,10*
0.43	
0.87	Nback (T1 vs.T2): -0,08
0.09	T1 (N-back vs. Quiz): -0,01 T2 (N-back vs. Quiz): -0,02

Table 3. Multilevel analysis of the training data (n-back task performance as a dependent variable). Models with training sessions (time) only (MODEL 1) and training sessions plus initial working memory capacity level as predictor and moderator (MODEL 2).

	MODEL 1 Time only		MODEL 2 + Initial OSPAN score	
	b	(s.e.)	b	(s.e.)
Fixed effects				
Intercept	2.165	(.088)***	1.928	(.143)***
Time - linear (centered at 1st day)	.031	(.005)***	.016	(.007)*
Initial OSPAN score ~ (high / low)	---	---	.038	(.183)*
Time × Initial OSPAN score	---	---	.026	(.008)**
Random effects ([co-]variances)				
Level 2 (between-person)				
Intercept	.285	(.073)***	.286	(.075)***
Time - linear	.001	(.001)***	.001	(.001)***
Intercept and time	.006	(.003)*	.004	(.002) °
Level 1 (within-person)				
Residual	.151	(.009)***	.149	(.009)***
Autocorrelation	.339	(.039)***	.328	(.040)***
Model fit				
−2 log likelihood (χ2)	1069.32		1046.37	
Akaike's Information Criterion (AIC)	1083.32		1056.37	

Results from multilevel modeling. Unstandardized regression coefficients with standard errors in parer
°p=0.1, *p<.05, **p<.01, ***p<.001; All p-values are two-tailed
~ the predictor is dichotomous

ıthesis.

Table of Materials

Name of	Company	Comments/Description
GEx	n/a	authorial online platform:
IBM SPSS Statistics 26.0	IBM Corporation	SPSS software was used to compute
Inquisit version 4.0.8.0	Millisecond	software: tool for designing and
MATLAB R2018b	The MathWorks,	MATLAB software was used to compute
PsychoPy version 2 v.1.83.04	Jonathan Peirce;	open-source software
Sublime Text (version 2.0.2)	n/a	open-source software: HTML editor

RE: **JoVE60804R1 Working memory training impact on cognition in the elderly population**

Dear Editors:

We greatly appreciate the opportunity to send you a revised version of our manuscript. You can find it attached to this letter.

We would like to thank the reviewers and editor for their thoughtful comments and suggestions. We have revised the manuscript accordingly.

Point-to-point responses to the reviewers' and editor's comments are provided in the following pages.

Sincerely on behalf of all authors,

Olga Matysiak and Aneta Brzezicka

Point-by-Point reply.

Color code: original, our reply.

Editorial comments:
Changes to be made by the author(s):
1.Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

We proofread and corrected language of the whole manuscript.

2.JoVE does publish methods and techniques that have been used to generate data previously published in results-driven journals. In all cases, JoVE requires original manuscript text that does not overlap with any other article, either previously published or at any stage in the publication process. If an author uses representative results or figures from a prior publication, the author must obtain permission from the original publisher and include appropriate citation.

We added appropriate citation)at the end of the Introductory part), we have also asked for permission form the publisher and we were assured that we can re-use representative results. Below is the response from the publisher:

"Dear Dr. Brzezicka,

Thank you for your message. At Frontiers, the entire content of all present and past journals is immediately and permanently accessible online free of charge and published under the CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and the source are credited.

As you can see at the bottom of the article page, "*This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*".

I hope this helps and should you have any further questions, I will be happy to answer.

Very best wishes,
Melissa
--
Melissa Bill
Review Operations Assistant"

3. Please revise lines 206-209, 211-212, 217-220, 222-223, 238-240, 278-286, 346-349, 363-365, 366-370, 373-376, 386-390, 401-403 to avoid textual overlap with previously published work.

The line were rewritten or changed.

4. Title: Please rephrase it to reflect the content of the protocol.

The title has been rephrased.

5. Please break the Introduction into several paragraphs describing the following: A clear statement of the purpose of this method; The rationale behind the development and/or use of this method; The advantages over alternative methods with references to relevant studies; The context of the method in the wider body of published literature; Information to help readers decide whether the method described is appropriate for them.

We modified introduction in order to reflect all suggested issues.

6. All methods that involve the use of human or vertebrate subjects and/or tissue sampling must include an ethics statement. Please provide an ethics statement at the beginning of the protocol section indicating that the protocol follows the guidelines of your institution.

We have added the information about ethical aspects of our study together with adding recommendations to the protocol description.

7. Please revise the Protocol text to avoid the use of personal pronouns (e.g., I, you, your, we, our) or colloquial phrases.

We corrected the language throughout the Protocol.

8. Please use the active/imperative voice and complete sentences throughout the protocol. Any text that cannot be written in the imperative tense may be added as a "NOTE."

We corrected the language throughout the Protocol.

9. Please describe how to carry out the Mini Mental State Examination. Alternatively, provide a relevant reference here.

We added the information about MMSE.

10. Please organize the sections/steps properly so that the protocol can be followed in chronological order.

We corrected the Protocol accordingly.

11. After you have made all the recommended changes to your protocol section (listed above), please highlight in yellow up to 2.75 pages (no less than 1 page) of protocol text (including headers and spacing) to be featured in the video. Bear in mind the goal of the protocol and highlight the critical steps to be filmed. Our scriptwriters will derive the video script directly from the highlighted text.
12. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. The highlighted text must include at least one action that is written in the imperative voice per step. Notes cannot usually be filmed and should be excluded from the highlighting.

We followed these recommendations during correcting the Protocol.

13. JoVE is a methods-based journal. Thus, the discussion section of the article should be focused on the protocol and not on the representative results. Please revise the Discussion to explicitly cover the following in detail in 3-6 paragraphs with citations:
a) Critical steps in the protocol
b) Modifications and troubleshooting of the method
c) Limitations of the method
d) The significance of the method with respect to existing/alternative methods
e) Future applications or directions of the method

We have changed the discussion part of the article in order to include all abovementioned points.

14. Please upload each Table individually to your Editorial Manager account as an .xlsx file. Avoid any coloring or formatting in the tables.

We prepared all tables in .xlsx format.

Reviewers' comments:
Reviewer #1:
Manuscript Summary:
The authors present a manuscript describing effects of a working memory on the cognitive functions in older age. They used the n-back task as a training intervention and a Quiz as an activity of the control group. This is an important issue and there is still no consensus about efficacy of cognitive training.

Major Concerns:
This study sounds solid, the methods, tasks the randomization methods are correct. However, the most serious problem is that exact the same data were presented previously (Matysiak O, Kroemeke A and Brzezicka A (2019) Working Memory Capacity as a Predictor of Cognitive Training Efficacy in the Elderly Population. Front. Aging Neurosci. 11:126. doi: 10.3389/fnagi.2019.00126). Unfortunately, the authors did not consider or cited their previous work indicating self-plagiarism and violation of ethical standards.
Nevertheless, as I don't know the policy of the journal in detail I read the manuscript and invested some time to give the authors some suggestions to improve their future studies.

Crucial function in the n-back task is not only the "working memory" concept but a complex interaction of several basic functions like updating (online update of memory contents), maintenance (simultaneous maintenance of contents, like in the dual task) as well as interference processing (resistance against distracting stimuli, i.e. non-targets). The proposed test battery does not cover all these important functions underlying the n-back task.

We agree that our battery do not include all possible functions involved in n-back task. Nevertheless we tried to use representative tasks from different levels of complexity (e.g. simple STM tasks and complex linear syllogisms) in order to be able to detect any form of transfer.

The largest problem of the training is the motivation to train the relatively boring task across 25 sessions. The task doesn't offer any changes like for example involvement of some less demanding training units, or relaxation games. How do the authors want to motivate the participants to do this?

An important prerequisite to learn is feedback. Do the participants get some positive or negative feedback about the progress of the learning process?

Our participants did receive the information about their progress after each session. All of our participants were older adults and all of them were very motivated to participate in training study. We did not have a big drop-out from the study, what is another indication that our participants were motivated. We did use reminders if we noticed that somebody did not train for more than two days.

The Quiz selected as control activity is perhaps more interesting and less boring activity that may even increase the motivation to train regularly. When the authors really look for an appropriate control activity this should contain the same elements as the n-back training and conduct for example interchangeably the visual and auditive task in a 1-back format. Such a version would equal the sensory input and motivation to train and would make the groups comparable.

We specifically did not want to use a 1-back format as a control task. We used it in our preliminary studies and it was a big mistake as our participants explicitly complained about this tasks and we had a huge drop-out (about as much as 80% of our participants did not finish the training). Quiz task was selected after careful consideration of what is at least as attractive as a regular n-back task (not less).

It is unclear why the authors don't present all results and the corresponding statistics. Unfortunately, none of the interactions Session x Group reached significance (presumably due to factors listed above). Min effect of Session in the syllogism task may be merely due to repeated application of the test and p-value of .09 for the interaction in the Go/NoGo task is far from significance. Thus, the conclusions that "comparing the effectiveness of both trainings showed differences (in favor of the experimental - WM training) only in the tasks engaging cognitive inhibition and WMC" is not justified.

We have changed our phrasing to not overestimate the WM training effectiveness. Nevertheless, we based our interpretation on the fact that we had statistically significant simple effects (pre-post) favoring WM training group.
Abovementioned effect of interaction between training group and time for go / no go task is not statistically significant, however as presented in Table 2, can be treated as statistical tendency: $F(1,83)=2,82$ $p = 0,10$. We corrected our descriptions of those effects throughout the text, including abstract.

The MLM is not a usual method to model experimental data. Please describe in more detail whether this method is based on linear mixed models using maximum likelihood estimation.

As written in the text, a method used to calculate predictors of training effectiveness is based on linear mixed models using maximum likelihood estimation. This approach is used in analysis of a data driven from experimental research (e.g. Bolger, N., & Laurenceau, J.-P. (2013). Intensive Longitudinal Methods:An Introduction to Diary and Experience Sampling Research. New York:Guilford Press).

Table 1 shows results of the MANOVA of the test battery. However, the table presents only one dependent variable per task instead of two for reaction times and errors. Individual standard deviations (ISD) or another parameter describing individual variability as a dependent variable would also be useful. Moreover, descriptive statistics (means and SDs) are important to include the study potentially in a meta-analysis.

The lack of this data is our oversight. Mentioned statistics were added to the manuscript in the form of a table (Table 1). Considering the comprehensiveness of other analyzes (MLM models) we decided to limit the analysis to one indicator per test. Especially that preliminary analysis revealed that results based on reaction times showed the significance of the same relationships between variables.

Minor Concerns:
Participants' recruitment
It is unusual to cooperate with professional companies to recruit volunteers. What is the advantage of such company and what are the methods to acquire the participants?

The decisive factor for employing professional recruitment companies in a country such as Poland, in addition to the very clear saving of time of the research team, is primarily the possibility of devolving to this company restrictive legal obligations (GDPR) related to the collection and operation of personal data of applicants.

Elderly participants who use social media or announcements at Universities of a Third Age are a highly selected group (well educated, motivated, interested in research and well financially situated). Several studies showed larger cognitive resiliency and reserve in these groups that may in turn substantially affect the training effects. Consequently, the conclusions based on training effects cannot be generalized to the general population. Therefore, it would be recommended to reach older adults with low educational and financial status.

Volunteers were recruited in the largest city of Poland, not only in through online channels, but also in person at various social events. The fact is that elderlies taking part in this type of activity are motivated and more pro-active than seniors who for example do not leave their homes. However, the level of education and economic status (indirectly controlled - as a occupational activity that generate income) were measured in the study. The analysis showed that these were not factors affecting training progress; therefore their role is not further analyzed in this work. We added additional sentences to discussion part covering this issue.

The switching task is not sufficiently defined, for example the type of paradigm (alternating runs (AABB) or random with or without preparation, overlapping stimulus-response assignments), presentation times, the length of response-stimulus intervals etc.

Missing information has been completed in the text.

Training protocols:
Why the access was restricted to "every 24 hours"?

Participants had limited access to the platform so that they were not able to perform more than one workout a day.

Reviewer #2:
Manuscript Summary:
Research on the effectiveness of cognitive function training still remains an important issue of psychological research, both in terms of knowledge development and application value. At the same time, despite many studies, the results are still inconsistent and there is no consensus on the conclusions. One of the reasons for obtaining different results is the use of different research protocols. In this context, the presentation of research protocols is very

important. The article is interesting and important, but it contains some discrepancies which should be clarified before publication.

Major Concerns:
There is a discrepancy between the information on the sample size and the age of study participants given in the abstract and in the text of the article. In the abstract is the information that the n-back training was conducted in the experimental group of 85 participants (30 males) aged 55-81 (lines 38-39). In the study protocol in turn a minimum age of 60 at is indicated as a criterion for inclusion to the sample (line 89).

Moreover, in the paragraph on the results there is information that the whole sample (not only experimental group) consisted of 84 persons (29 males), and these 84 people were divided into experimental and control groups. Thus, the experimental group that trained with n-back task included 42 participants, not 85, as stated in the abstract.
If there were actually people aged 55-59 in the sample, it is necessary to justify their inclusion to the category of "older adults".

This information has been corrected.

It is important to provide information what did the protocol provide for when the screening tests indicated the presence of cognitive disorders (apart from exclusion from the sample) - were the participants informed of the result?

Volunteers in the recruitment process were asked whether in their history they had diagnosed neurological diseases or cognitive impairments. If the answer was affirmative, such persons were not invited to participate in the study.
Information on what to do if the subject does not reach 27 points in the MMSE scale was added to the text.

The inclusion and exclusion criteria should be explicitly mentioned. They are currently given in several places in the text ('Participants recruitment' and 'Initial screening'). It is not clear whether all the criteria are listed, e.g. are sight, hearing and motor skills taken into account?

As recommended, the criteria for excluding people from the study were collected and described in one section ‘Participants recruitment'.

It is not clear what is the authors' idea about the status of the active control group - sometimes it is treated as a typical control group (in which no improvement should be observed, as opposed to an experimental group), sometimes as a second, alternative, training group. What was the basis for the decision to train semantic memory in the active control group? What were the hypotheses about the effects of semantic memory training/quiz and what were the assumptions and premises for these hypotheses?

We did not have specific hypotheses regarding our control group, nevertheless we did observe similar improvement in our cognitive battery. Unfortunately we did not register the training data at such precise level as for our training (n-back) group, so we were not able to perform similar to n-back group analysis (MLM) on Quiz group.

Describing "Training group assignment" the Authors stated "Make sure that your group assignment is not biased in terms of age, gender or the level of initial WM capacity". However, they do not provide the data confirming that they have taken this principle into account: comparisons of experimental and control groups in terms of age, gender, and baseline WM capacity. It seems to be the opposite. The results presented in Table 1 for the Ospan task indicated that there is no significant effect of the group and time interaction. In turn, the group effect is statistically significant. It suggests that already in baseline there were differences between groups in terms of WM capacity.

We added this recommendation specifically because of our main discovery – initial WM or more broadly cognitive level being a strongest predictor of training effectiveness. As our analysis with WM as a training effectiveness predictor was based on n-back training group, the differences between groups were not that relevant. Moreover, the initial WM variability – in the case of our analysis – probably enabled us observation of this effect.

As far as Table 1 is concerned, I have doubts whether it actually shows the results of MANOVA. MANOVA is used for comparing multivariate sample means, when we have more than one depended variable. In Table 1 are presented the results for each individual dependent variable. Maybe these are the results of one-way tests which supplemented MANOVA, or a series of RM ANOVA with group as a between-subject variable?
Data presented in the Table 2 (Apart from (or instead) the difference between the means (in Table 1), it is worthwhile to show the means themselves - the size of the difference between the means has a different meaning depending on the values of the means (e.g. difference 1.5 for means 1 and 2.5 and for means 12 and 13.5), moreover, the results of different tests are expressed on different scales.

The data presented in Table 2 (formerly named Table 1) show the result of one analysis performed using the SPSS package and functions: GLM - repeated measures, where all test's results were dependent variables while the time and group affiliation were independent variables.
Abovementioned statistics were added to the manuscript in the form of a table (Table 1).

I also have doubts about the interpretation of some of the results presented in Table 1. According to the results, there are no significant effects of interaction, in other words, in the experimental group there was no improvement (or change) greater (or different) than in the control group. In this context it is questionable to conclude that n-back training is effective. At the same time, the interpretation of these results would have been clearer if the explicit assumptions about the role of the active control group would be formulated and if the results anticipated in these group would be clearly indicated beforehand.

We have changed our discussion part to incorporate this issue.

Minor Concerns:
The title suggests that the research was of a population-based nature, which was not the case (the sample included 84 participants). In my opinion the term " in older adults" would be more relevant.

The title was changed.

It follows from the text that the participants did not have to agree to initial screening tests. What justifies the lack of such consent?

The participants gave the consent before starting the initial screening. The description of this process has been improved taking into account the above remark.

In the section "WMC as a predictor of WM training effectiveness" the Authors have written on 3 models (lines 302-304). However, in the Table 2 are presented results regarding only 2 models. What is the reason of that?

The number 3 was used by mistake instead of 2. The change was done in the text.

And in the case of the interaction effect "Time x Initial OSPAN score", isn't baseline OSPAN score treated as a moderator, not a predictor?

Model 2 described in section ‘WMC as a predictor of WM training effectiveness’ beside verifying the time as a predictor of number of N change, added on predicting as well as moderating effects of a baseline OSPAN score (between-person predictors - level 2) on within-subject variation (moderation - level 1). This information was included in the text.

RE: **JoVE60804R2 "The importance of initial intellectual functioning assessment in working memory training effectiveness evaluation in older adults".**

Dear Editors:

We greatly appreciate the opportunity to send you a revised version of our manuscript. You can find it attached to this letter.

We would like to thank the reviewers and editor for their thoughtful comments and suggestions. We have revised the manuscript accordingly.

Point-to-point responses to the reviewers' and editor's comments are provided in the following pages.

Sincerely on behalf of all authors,

Olga Matysiak and Aneta Brzezicka

Point-by-Point reply.

Color code: original, our reply.

Editorial comments:
Changes to be made by the author(s):
1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

We proofread and corrected language of the whole manuscript.

2. Please revise lines 133-135, 232-235, 252-255, 257-258, 263-269, 283-286, 328-334, 406-409, 413-416, 419-421, and 425-427 to avoid textual overlap with previously published work.

The lines were rewritten or changed.

3. Title: Please rephrase it to reflect the content of the protocol.

The title has been changed to reflect the content of the protocol.

Old title: The importance of initial intellectual functioning assessment in working memory training effectiveness evaluation in older adults

New title: Working memory training for older participants – the importance of the control group's training regimen and initial intellectual functioning assessment.

4. Please revise the Protocol text to avoid the use of personal pronouns (e.g., I, you, your, we, our) or colloquial phrases.

We did our best to correct language in order to avoid using personal pronouns and colloquial phrases.

5. 3.4.2: How to perform an assessment of orientation to place and time?

7. 4.1: What detailed instruction is given?

We added fragment on assessment of orientation to place and time.

Reviewers' comments:
Reviewer #2:
Manuscript Summary:
After the corrections made by the authors I have no further comments or suggestions for changes in the article.

Minor Concerns:
I still think that the results referred to as MANOVA results are the results of one-way tests which supplemented MANOVA. MANOVA statistics are given for all dependent variables at once, not separately (and we conclude that the respective effect is significanthe if overall multivariate test is significant). The most common test statistics for MANOVA are Wilks' Lambda, Pillai's Trace, Hotelling's Trace or Roy's Largest Root, but they are not referred in

the article. I therefore assume that MANOVA and one-way supplementary tests were performed, but only the results of the latter were shown.

We corrected description.

Reviewer #3:

We are grateful for Reviewer #3 comments; they were very helpful in improving our manuscript.

Major Concerns:
- Introduction and discussion should be shortened regarding the discussion of the implementation of the control training. On the one hand, the authors claim that it is a major scope of the article to compare two cognitive interventions and they explain the implementation of the control group in detail. On the other hand, the choice of the specific control group appears to be a minor issue due to the missing group differences and the focus of the individual factors. That is why I suggest to include the reason of choosing a memory quiz as control regimen and its implementation in the respective section in the protocol (p. 6, line 300) but to remove or shorten it in the rest of the article (e.g., p.1 line 79-83, p. 9 line 424-439, p. 10 line 460-462) in favor of a clearer statement regarding the major aim of the study - the analysis of working memory training efficacy and the impact of the initial working memory capacity.

We introduced suggested changes to the manuscript.

- Please provide how you determined your sample size - was a power analysis conducted prior to data collection? This should be included in section 1.1. (p. 2 line 95-99) - to collect as many participants as fast as possible is not an ethically correct approach.

We provided additional information on sample size setting.

- Results regarding training-related effects: Please provide the specific statistical analyses you conducted to produce the results reported in the text and table 2 (e.g., statistical tests used, effect size reported and its interpretation);

We added this information to the text.

Please state whether significant differences were found between the groups at baseline (the raw data provided in table 1 suggests that the performance differed in the OSPAN and the memory SPAN task);

We added this information to the text.

Do not report results that are "close to significance"! Either the statistical significance is given or not, there is no such thing as "almost significant" or "by trend"! Please correct all your statements in the results and discussions section in consideration of this fact (e.g., p. 7 line 331: NO significant interaction was found regarding the training regimen).

We corrected the results' section accordingly.

The "training group effects" (independently of the time point of assessment) reflect that the training groups differed from the beginning which makes it difficult to compare their cognitive outcomes, this should be reported.

We reported it in the manuscript.

In the end, the results section should be presented more clear:
a) in the trained tasks only improvements in the n-back group (partly due to adaptivity), worse performance in the memory quiz group (can you provide more specific data here, eg. percent correct answers pre-post),
b) existing pretest/group differences (if so) in the baseline measure of... ,
c) unspecific significant improvements of both training groups in the transfer measures of... ,
d) no significant interactions.

We corrected our description of the results section.

- Results regarding WMC as predictor of training efficacy: please state more clear in the beginning that this section only refers to the analysis of the n-back training group;

We added this information.

Why did you pick the OSPAN as criterion measure and not the other working memory tasks - or even better - a latent variable of the construct or a composite score? Do the other variable predict training efficacy as well and if not, why?

This is a very good point. In fact, we tried both suggested methods: calculating composite score and latent variable. However in case of factor analysis there were no clear latent variables found. We believe that composite score as a variable in MLM model is not the perfect choice – one of our goals was finding factors that influence learning curve. Making them as simple as possible has not only statistical benefits, but also interpretational – using simple variables makes practical implementation of the results less complicated.

Have you tried to predict the performance in the transfer tasks by individual working memory capacity (please report whether or not the prediction is restricted only to the trained tasks and if so, provide argumentations why high performers do not seem to benefit on the transfer outcomes)?

For the purpose of this article this analysis was not performed. This article is focused on protocol, so we did not perform additional analysis.

- Instead of the limitations presented (p. 10 line 451-456 regarding the selection of high performing elderly adults), please consider including the following important methodological aspects in the limitations section: no adaptivity of difficulty in the memory quiz group may have resulted in lower motivation towards the end of the training phase compared to adaptive working memory training (this may be indicated by a lower performance in the final session as in the first session, see p. 7 line 334);

Training progress in memory quiz wasn't measured, therefore we do not have the data regarding changes 1st vs. last session performance in the memory quiz group.

Minor Concerns:
We corrected all issues mentioned by the reviewer

Reviewer #4:
Manuscript Summary:
The present manuscript details an experiment designed to elucidate the role of individual differences in baseline WMC on training and transfer outcomes following a five week cognitive training regimen. The experimenters found that baseline indeed predicts performance on the intervention, and that those with higher OSPAN performance at baseline had a steeper learning curve than those with lower performance on the untrained measure.

Overall, there is a strong need for additional work that considers the effects of individual differences in the outcome of cognitive training, and particularly the role of baseline performance (see Rhodes & Katz, 2017, and Katz, Jones, Shah, Bushkuehl, and Jaeggi, 2016), especially among older adult populations.

The article is about **the protocol and do not** present new results. The results of this study were already published so we are not responding to this reviewer's concerns.