Practice effects in nutrition intervention studies with repeated cognitive testing

Article

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Research Report

Practice effects in nutrition intervention studies with repeated cognitive testing

Lynne Bell, Daniel J. Lamport, David T. Field, Laurie T. Butler and Claire M. Williams*
School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Whiteknights, Reading, UK

Abstract.
BACKGROUND: There is growing interest in the use of nutrition interventions to improve cognitive function. To determine intervention efficacy, repeated cognitive testing is often required. However, performance on tasks can improve through practice, irrespective of any intervention.
OBJECTIVE: This study investigated practice effects for commonly used cognitive tasks (immediate and delayed recall, serial subtractions, Stroop and the Sternberg task) to identify appropriate methodology for minimising their impact on nutrition intervention outcomes.
METHODS: Twenty-nine healthy young adults completed six repetitions of the cognitive battery (two sessions on each of three separate visits). Subjective measures of mood, motivation and task difficulty were also recorded at each repetition.
RESULTS: Significant practice effects were apparent for all tasks investigated and were attenuated, but not fully eliminated, at later visits compared with the earlier visits. Motivation predicted cognitive performance for the tasks rated most difficult by participants (serial 7s, immediate and delayed recall). While increases in mental fatigue and corresponding decreases in positive mood were observed between test sessions occurring on the same day, there were no negative consequences of long term testing on mood across the duration of the study.
CONCLUSIONS: Practice effects were evident for all investigated cognitive tasks, with strongest effects apparent between visits one and two. Methodological recommendations to reduce the impact of practice on the statistical power of future intervention studies have been made, including the use of alternate task forms at each repetition and the provision of a familiarisation visit on a separate day prior to data collection.

Keywords: Cognition, practice effects, nutrition intervention, methodology

1. Introduction

Diet is well known to have an impact on human health with many recent studies suggesting that nutrients present in natural foods can have a positive effect on our mental function as well as physical health e.g. [1–6]. There is growing interest in the use of dietary interventions to promote lifelong cognitive health, particularly with a view to reducing the incidence, or delaying the onset, of age-related neurodegenerative diseases e.g. [4, 7–9].

Repeated cognitive testing is often necessary in order to determine the efficacy of nutrition interventions over time, particularly when determining a dose response curve. Equally this situation may arise in clinical drug trials or with other intervention types where repeated cognitive testing occurs over time. In such studies, multiple time points are used for comparing test values with baseline values and, in either acute or chronic crossover studies, or those utilising parallel group designs, testing may be performed multiple times by the same participant over a period of many weeks or months. However, the effects of repeatedly practising cognitive tasks are known to be problematic for studies using a repeated-measures design [10–17]. Each time a participant is asked to perform a task they become more familiar

*Corresponding author: Professor Claire M. Williams, School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Whiteknights, Reading, RG6 6AL, UK. Tel.: +44 0118 378 7540; Fax: +44 0118 378 6715; E-mail: claire.williams@reading.ac.uk.

with both the procedure and stimuli presented, and a process of learning takes place, often leading to enhanced performance. This becomes increasingly problematic if performance approaches or reaches a ceiling. Additionally, practice may result in a change in the participant's strategy for performing the task [18], which may in turn modify the brain network being utilised e.g. [19, 20], calling into question the validity of the result. Practice effects also add additional error variance that may impact on the statistical power of the study [21]. In order to ensure the validity of cognitive research, the European Food Safety Authority (EFSA) guidelines on psychological health claims currently specify that practice effects must be addressed in behavioural intervention studies with a repeated testing component [22]. At least one methodological review paper [23] has identified practice to be a problem for cognitive tasks across a number of cognitive domains. Many research groups adopt a range of strategies for dealing with practice effects; however there is no current consensus on the best methodology to address this issue.

Conceptually, very few cognitive tasks were developed with repeated-measures testing specifically in mind. Instead of seeking practice-resistant tasks, which are likely to be difficult or even impossible to create due to human adaptive behaviour, it has become almost standard practice for studies using multiple testing points to adopt the use of alternative forms of a task. Using this method, the same task is used but different equivalent forms of stimuli are presented across the multiple testing points. However although this strategy has previously been shown to be effective at attenuating practice effects, for some tasks significant residual practice effects may still be evident [24]; it is thought that for many tasks participants are able to develop strategies to enhance their performance over time irrespective of the specific stimuli presented. The use of alternate forms cannot fully counteract this procedural learning process [25]. A number of studies attempt to reduce practice effects by incorporating an additional task familiarisation session before the test sessions, where participants familiarise themselves with the tasks, either on a prior visit or immediately before data collection, with a view to raising performance to a more stable level before beginning data collection. Indeed, this technique was historically advocated by McClelland [17], who recommended a minimum of four familiarisation sessions prior to data collection for some tasks, although it is unclear from the paper whether these four sessions should be spread across one or more

visits. The addition of this number of sessions has significant time and cost implications and has not been universally adopted due to its impracticality. However, many studies include upwards of one familiarisation session, with some including up to four or five, at a prior screening visit e.g. [26–31], although it has been acknowledged that others entirely forgo adequate training [32]. The effectiveness of different strategies, including the most appropriate time for conducting familiarisation sessions (separate visit or immediately before testing) have not been fully investigated.

The effect of practice on performance has previously been investigated for short term repeated testing on a single visit e.g. [10, 13], where participants performed four repetitions of a battery of tasks at 10 minute intervals, and for repeated testing over a longer term with single performances of a task across multiple visits e.g. [14–16, 21, 24, 33]. A few studies have combined within visit and multiple visit sessions of testing e.g. [10, 17]; however in the latter example the majority of tasks were conducted using pencil and paper and were observed to be relatively unaffected by practice. Combined, the evidence from the above studies suggests that significant practice effects are observed for most modern computerised tasks, with the strongest practice effects typically observed between the first and second testing time points irrespective of their timing [10, 13, 16], although significant practice-related improvements have been observed beyond a second session of testing [16, 24, 33]. Practice effects have also been observed both within individual tasks, and within a test battery depending on the temporal positioning of individual tasks [13]. Overall there is some suggestion that the rate of practice-related improvement may slow after two or more sessions irrespective of the testing interval. However, the timing of the repeated sessions used by these studies (e.g. minutes [10, 13], weeks [16, 21, 24, 33], or months [14–16, 21]) generally do not reflect those typically used in crossover nutrition studies or similar clinical intervention trials, where it is common to measure performance at baseline and again at one or more post- intervention time points at each visit, over a number of weekly visits e.g. [28–31, 34–36]. This is important as practice effects have been observed to differ depending on the testing interval [37]. Similarly, in the parallel field of learning and memory, the timing of practice sessions is known to be important, as distributed practice spread across days, weeks or months has been observed to facilitate differential learning compared with

massed practice performed multiple times on a single occasion [38].

This study aimed to investigate practice effects for a number of commonly used cognitive tasks within the design framework of a crossover nutritional intervention; crossover designs are likely to incorporate a larger repeated testing component than parallel designs as participants are actively involved in all control and intervention arms of the study. In the current study three test days and two testing points within each day (a 3 × 2 design) were selected. The chosen design was broadly representative of those used in this field e.g. [28–31, 34–36], although the overall number of test days and testing points may vary from study to study. This design could also be equated to a single arm of a parallel design study with multiple test points. In line with current standard practice, alternate forms of all cognitive tasks were used, and familiarisation trials immediately prior to data collection were also incorporated in the study design. The intention was to better understand practice effects in the context of an intervention study design with a repeated testing component, and specifically to identify the extent of practice effects within a cognitive battery which has subsequently been applied to a series of nutritional interventions over the course of a doctoral research degree. The data were used to determine a practical yet effective strategy for minimising and accounting for practice effects and the associated variance in performance in nutrition intervention studies, as recommended by EFSA.

2. Materials and methods

2.1. Ethical standards

This study was approved by the University of Reading ethics committee and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All persons gave their informed consent prior to their inclusion in the study.

2.2. Participants

The participants were 29 apparently healthy young adults, aged 18–42 years (mean 25.6, SD 7.7, 9 male). Following a power analysis, 30 participants were recruited via email and social media from staff and student populations at the University of Reading;
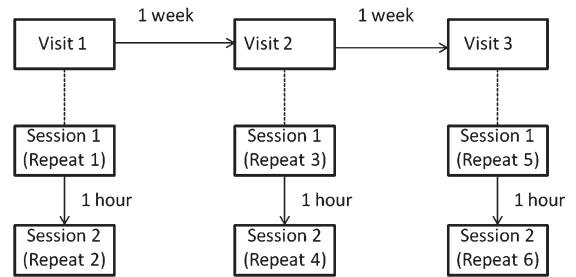

Fig. 1. Study design; the study was structured to mimic a crossover design with three different interventions tested at baseline and 1 hour post-intervention, across three weekly visits. Repeat numbers refer to successive exposures to the cognitive task battery, with Repeat 1 being the first attempt.

however one participant failed to attend any test visits. Subjects were non-smokers, but no other exclusion criteria were applied.

2.3. Design

The study design is illustrated in Fig. 1. Each participant completed the cognitive test battery six times following a crossover design. The six test sessions were split over three visits, with each visit separated by approximately one week (mean 7.02 days, sd 1.12 days, range 3–11 days). Within each visit there were two test sessions which were separated by 1 hour. The cognitive battery lasted 40 minutes. Participants were tested at the same time of day on each visit to minimise diurnal effects. The study was reviewed by the University of Reading Ethics Committee and given approval to proceed.

2.4. Procedure

Participants attended the lab for a total of three visits. On arrival they immediately completed the battery of cognitive tasks and subjective measures of mood outlined below. Participants then waited in the lab for one hour before repeating the testing. During the break participants were supplied with magazines to read. While no specific dietary restrictions were in place, participants tested in the morning ($n = 15$) were asked to record their breakfast intake including all beverages (caffeinated or otherwise) on the first test visit, and were asked to eat the same breakfast at the same time prior to all subsequent visits. Participants tested in the afternoon ($n = 14$) were similarly asked to eat the same lunch including all beverages at the same time before all visits. During all visits

only consumption of water was permitted. After testing had been completed a return appointment was arranged for the following week. Participants spent a total of 2 ½ hours in the lab at each visit. Participants were informed from the outset that the aim of the study was to investigate the effects of repeated practice on cognitive task performance.

2.5. Cognitive measures

The tasks were broadly representative of three main cognitive domains: working memory (serial 3s and 7s subtraction; Sternberg memory scanning), executive function (Stroop) and episodic memory (immediate and delayed recall), all tasks also required attentional processing. The selected tasks have all been commonly used in previous crossover nutrition intervention studies [12, 39, 40]. The tasks were all programmed using E-Prime. With the exception of the immediate and delayed recall tasks, which were respectively presented first and last, the order of the remaining cognitive tasks was counterbalanced between participants. Serial subtraction, 3s and 7s, were included as a single task in the counterbalancing, with serial 3s always performed before serial 7s. Therefore, there were six unique orders, all of which were repeated. Short practice trials were incorporated at the start of each of the executive function and working memory tasks, for which data were not collected. This was to ensure that participants had fully familiarised themselves with the task at each administration, so within-task practice effects were minimised [13]. Alternate forms were used for all tasks.

1) Immediate and delayed recall. This episodic memory task was a single-trial word list learning task with immediate and delayed free recall components. Fifteen words were visually presented in sequence on a computer screen at a rate of one word every 2500 ms. Each word remained on screen for 2000 ms. At the end of the presentation participants were asked to recall as many of the words as possible (no time limit), writing them on a piece of paper (immediate recall). This task was the first task administered during each test block. After all other tasks had been completed, participants were again asked to recall the words (delayed recall; approximately 30 mins post initial presentation). Different word lists, matched for word length, concreteness and familiarity, were used at each test session [41]. The dependent variable was the total number of correct words recalled.

2) Word-Colour Stroop. This was a modified, computerised version of the original executive function task [42]. The words "Purple", "Green", "Blue", and "Red" were displayed individually in a randomised, counterbalanced order, with each word being displayed in either a congruent or an incongruent ink colour. The words were presented at a rate of one every 2000 ms and each word remained on screen for 1250 ms. Participants were instructed to respond to the ink colour as quickly as possible by pressing the corresponding coloured button on the keyboard (coloured stickers were placed over the numerical keys 1–4). Twelve practice trials and 96 test trials were presented at each test session. An equal number of congruent and incongruent trials were presented. The interference effect of the semantic meaning of the word was calculated by subtracting the mean reaction times (for correct responses only) for congruent trials from incongruent trials. Responses slower than 1250 ms were recorded as errors; however in the young adults tested, accuracy rates were consistently high with a number of participants performing at ceiling and so accuracy has not been reported as a dependent variable here.

3) Sternberg memory scanning. This was a modified, computerised version of the original task developed by Sternberg [43, 44]. The task measures how fast participants can scan through a list of items held in their short term memory. Previous research has shown this is a fixed time per item; therefore the task has been described as resistant to practice [45–47], although it should be noted that underlying reaction times from which this measure is derived are still subject to practice related improvements. During the Sternberg task, participants were presented with a sequential series of 1 to 6 digits, the order of which they were required to memorise. This was a necessary component of the task in order to measure working memory rather than episodic recognition memory, and compliance was assessed by a memory test on random trials throughout the task. However, due to the random nature of the measure it was not included as a dependent variable. A new set of digits was randomly generated on each trial. The appearance of a fixation point indicated the end of the sequence. The digits and fixation were each presented for 1200 ms at a rate of one

every 1200 ms. Participants were required to indicate as quickly as possible, with a labelled yes/no key press ('b' and 'n' respectively), whether or not a final digit, presented 2000 ms after the fixation point, was present in the original memory set. The task was self paced with no time limit for responses. Participants completed 12 practice trials and 96 test trials at each test session. The dependent variables were mean scanning rate and mean RT. These variables were, respectively, the slope and intercept from the regression model for predicting reaction time from memory set size [48–51]. Accuracy rates were important in order to provide sufficient power for the regression analysis, therefore performance feedback was given after each trial. As with the Stroop task, accuracy was consistently high with a number of participants performing at ceiling and so has not been reported as a dependent variable here.

4) Serial Subtraction, 3s and 7s. Using a previously published method for this working memory task [52–54], a random number between 800 and 999 was presented on screen and participants counted backwards, at first in 3s, entering their answers via the computer number pad as quickly as possible for a total of 2 minutes. The task was then repeated subtracting 7s instead of 3s. The dependent variables in both cases were the total number of correct responses and mean RT (for correct responses only). The accuracy of the response was determined relative to the previous response, irrespective of whether or not the previous response was correct. A 20 second practice trial was completed immediately before each 2 minute test.

2.6. Subjective measures

Measures of task difficulty and motivation were recorded on paper using a nine point Likert scale questionnaire after completing each cognitive test. The wording on each scale was as follows: 'How difficult did you find the <task name> task?' and 'How motivated were you to do well during the <task name> task?' Anchor points were 1 'Not at all' and 9 'Extremely'. Therefore high scores indicated a high level of difficulty or a high level of motivation respectively. These measures were repeated each time a particular task was performed. In addition, a computerised version of the PANAS-Now [55] mood questionnaire with an additional subjective nine point Likert scale measure of mental fatigue [52] were

administered at the end of each session of cognitive tasks.

2.7. Data analysis

All data were analysed using IBM SPSS statistics version 22. Parametric tests were used throughout after confirming that test data met the required assumptions of normality. For all RT data, only correct responses were included in mean values.

The DVs for the Sternberg task were coefficients derived from a regression analysis predicting reaction time from the number of items scanned into short term memory as described in the original test. In order to preserve statistical power, the regression analysis was performed using all RT data points, not mean RTs.

A linear mixed model (LMM) using a first-order autoregressive heterogeneous covariance structure (ARH1) to model successive repeat test sessions was used to analyse data for all cognitive tasks. Visit, Session and the Visit*Session interaction were included as fixed factors in the model. Subjects were included as random effects. Motivation was included in the model as a repeated covariate (Model 1). Due to a lack of statistical power, time of day was not included as a factor. The analysis aim was to determine whether practice related improvements in cognitive performance were evident following the use of same day familiarisation trials and alternate forms. *Post hoc* comparisons were used to investigate any significant effects of Visit or Visit*Session. A Bonferroni correction was applied to all multiple comparisons. Cohen's d values were calculated to compare practice effect sizes between visits and sessions, with a view to determining whether a separate familiarisation visit, in addition to the use of same day familiarisation trials and alternate forms, would reduce practice effects. The same LMM procedure, without the motivation covariate (Model 2), was used to determine the effects of repeated testing on mood and mental fatigue.

Difficulty ratings recorded at each time point and for each cognitive task were analysed using a separate LMM (Model 3). A simple diagonal covariance structure was used. Cognitive task order was counterbalanced between participants, therefore ARH1 was considered inappropriate. Cognitive Task, and Task*Visit, Task*Session and Task*Visit*Session interactions were included as fixed factors in the model. Subjects were included as random effects.

3. Results

Following publication, data supporting the results reported in this paper will be openly available from the University of Reading Research Data Archive at http://researchdata.reading.ac.uk/.

Sternberg accuracy rates were high (mean 90.43 out of 96, sd = 4.97). Coefficient of determination values for the underlying regression analysis of the Sternberg data were low ($R^2 < 0.380$), but were similar to those observed in previous studies e.g. [56], and were typical of behavioural data.

3.1. LMM analysis of the changes in cognitive performance across all test sessions with motivation as a repeated covariate

Mean test scores with standard deviations recorded at each time point are presented in Table 1. The LMM model results are summarised in Table 2. For clarity, only significant fixed effects are reported in the text.

3.1.1. Motivation

As a repeated covariate in the model, Motivation significantly predicted IR score [$F_{(1,136.77)} = 13.56$, $p < 0.001$], DR score [$F_{(1,127.22)} = 7.64$, $p = 0.007$], serial 7 s score [$F_{(1,135.85)} = 19.05$, $p < 0.001$], and Stroop congruent RT [$F_{(1,140.24)} = 5.83$, $p = 0.017$], such that higher motivation was associated with better performance.

3.1.2. Visit

Fixed effects of Visit were significant for IR [$F_{(2,54.20)} = 16.82$, $p < 0.001$], serial 3 s score [$F_{(2,49.95)} = 70.48$, $p < 0.001$] and serial 3 s RT [$F_{(2,53.89)} = 35.30$, $p < 0.001$], serial 7 s score [$F_{(2,46.26)} = 32.86$, $p < 0.001$] and serial 7 s RT [$F_{(2,66.14)} = 11.43$, $p < 0.001$], Sternberg scanning rate [$F_{(2,60.89)} = 6.87$, $p = 0.002$], and both incongruent [$F_{(2,52.41)} = 13.62$, $p < 0.001$] and congruent [$F_{(2,43.04)} = 6.65$, $p = 0.003$] RTs on the Stroop task.

Post hoc tests showed improvements between Visits 1 and 2 for IR [$p = 0.004$], serial 3 s score [$p < 0.001$], serial 3 s RT [$p = 0.001$], serial 7 s score [$p < 0.001$], Sternberg scanning rate [$p = 0.010$], and Stroop incongruent RT [$p = 0.018$]. Improvements between Visits 2 and 3 were observed for IR [$p = 0.010$], serial 3s score [$p < 0.001$], serial 7s score [$p = 0.020$], and Stroop incongruent RT [$p = 0.029$]. Therefore, 8 of 11 DVs showed practice effects across test days with 6 DVs showing significant changes between Visit 1 and 2 but only 4 showing significant changes between Visit 2 and 3.

3.1.3. Session

Fixed effects of Session were significant for DR, serial 3s score and RT, serial 7s score and RT, Sternberg RT, Stroop incongruent and congruent RT, and Stroop interference effect. Significant improvements between Sessions 1 and 2 were observed for serial 3s score [$F_{(1,76.87)} = 42.98$, $p < 0.001$], serial 3s RT $F_{(1,86.05)} = 16.59$, $p < 0.001$], serial 7s score [$F_{(1,69.02)} = 10.74$, $p = 0.002$], serial 7s RT [$F_{(1,67.33)} = 17.04$, $p < 0.001$], Sternberg RT [$F_{(1,55.60)} = 9.37$, $p = 0.003$], Stroop incongruent RT [$F_{(1,57.19)} = 24.57$, $p < 0.001$], and Stroop congruent RT [$F_{(1,42.89)} = 50.08$, $p < 0.001$]. Conversely, significant decreases in performance between Sessions 1 and 2 were observed for DR [$F_{(1,71.43)} = 17.08$, $p < 0.001$], and Stroop interference effect [$F_{(1,61.19)} = 6.50$, $p = 0.013$]. Therefore, 9 of 11 DVs showed significant effects across same day test sessions, with 7 DVs showing improvement, and 2 DVs showing a decline in performance.

3.1.4. Visit*Session interactions

The significant interaction for IR [$F_{(2,81.91)} = 4.36$, $p = 0.016$] was explained by a significant increase in performance observed between Sessions 1 and 2 at Visit 1 [$p = 0.041$], but not at any other visits. For Serial 3s score [$F_{(2,76.20)} = 3.87$, $p = 0.025$] improvements were observed between Sessions 1 and 2 on Visit 1 [$p < 0.001$] and Visit 2 [$p < 0.001$], but not on Visit 3 [$p = 0.179$]. Similarly for Serial 7s RT [$F_{(2,73.47)} = 4.39$, $p = 0.016$] the significant interaction was explained by improvement in performance between Sessions 1and 2 at Visit 1 [$p = 0.001$] and Visit 2 [$p = 0.029$], but not at Visit 3 [$p = 0.503$]. A significant interaction for Stroop incongruent RT [$F_{(2,74.68)} = 8.70$, $p < 0.001$] was explained by improvements between Sessions 1 and 2 on Visit 1 only [$p < 0.001$]. For Stroop congruent RT [$F_{(2,83.05)} = 3.52$, $p = 0.034$], larger improvements were observed between Sessions 1 and 2 on Visit 1 [$p < 0.001$], although moderate improvements were still evident between Sessions on Visit 2 [$p = 0.016$] and Visit 3 [$p = 0.001$]. In all interaction cases, therefore, practice effects between Sessions 1 and 2 were attenuated or eliminated altogether at later visits.

Table 1
Means and standard deviations for all measures at each testing time point

| Test type | Visit 1 | | | | Visit 2 | | | | Visit 3 | | | |
| | Session 1 | | Session 2 | | Session 1 | | Session 2 | | Session 1 | | Session 2 | |
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cognitive:												
Immediate recall score (correct/15)	7.52	2.01	8.00	2.27	9.03	1.97	8.59	2.11	10.21	2.51	9.24	2.61
Delayed recall score (correct/15)	6.00	2.38	5.45	2.64	6.69	2.38	5.41	2.95	7.90	2.94	5.79	4.14
Serial 3s score (correct in 2 mins)	32.79	14.37	40.07	16.23	41.62	18.45	46.72	16.78	49.59	19.31	51.21	18.80
Serial 3s RT (ms)	3515.13	1457.62	3094.01	1412.59	2935.73	1165.08	2724.01	1503.01	2621.67	1229.22	2446.37	1120.99
Serial 7s score (correct in 2 mins)	20.28	11.59	22.45	10.56	23.97	13.68	25.10	12.66	26.72	13.76	28.10	13.89
Serial 7s RT (ms)	5913.03	2941.48	5086.12	2157.89	5252.54	3000.46	4484.32	1670.49	4601.17	2085.15	4505.79	2017.93
Sternberg scanning rate (ms/item)	37.67	15.17	35.59	26.13	26.98	18.15	28.42	14.62	27.05	21.30	25.55	21.65
Sternberg RT (ms)	668.81	189.31	595.49	180.10	616.15	169.75	583.18	153.24	575.58	245.04	561.33	162.40
Stroop incongruent RT (ms)	783.37	90.70	729.09	93.62	743.47	93.88	734.31	97.31	725.15	109.73	715.61	89.13
Stroop congruent RT (ms)	714.54	97.83	661.10	101.75	681.62	99.68	663.20	109.07	673.58	119.38	640.12	96.09
Stroop interference effect (ms)	68.84	46.16	67.99	48.15	61.85	51.41	71.11	44.20	51.56	47.76	75.48	44.19
Mood:												
Mental fatigue (rating/9)	5.31	1.91	6.10	1.74	5.48	1.70	6.00	2.04	4.79	2.19	5.03	2.04
Positive affect (score/50)	24.66	7.59	21.97	8.35	23.24	9.28	19.55	7.55	22.76	7.86	23.00	8.50
Negative affect (score/50)	13.34	3.87	12.72	4.10	12.86	4.56	12.76	5.12	12.72	5.65	12.38	2.88
Motivation rating (out of 9):												
Immediate recall rating	6.76	1.57	5.72	2.05	6.62	1.63	6.03	2.18	6.55	1.94	6.10	2.09
Delayed recall rating	6.24	1.77	5.97	2.13	6.00	2.00	5.83	2.05	6.31	2.09	6.14	2.33
Serial 3s rating	6.10	1.95	6.31	1.91	5.69	2.16	5.69	2.38	6.38	1.90	5.97	2.11
Serial 7s rating	6.03	2.13	5.76	2.12	5.38	2.23	5.45	2.35	5.86	2.13	5.79	2.38
Sternberg rating	5.45	1.97	5.45	2.16	5.14	2.00	5.05	2.10	5.72	2.25	5.48	2.15
Stroop rating	6.34	1.74	6.14	1.68	6.14	1.85	5.97	1.90	6.24	1.94	6.07	2.17
Difficulty rating (out of 9):												
Immediate recall rating	6.14	1.60	5.97	1.43	5.31	1.56	5.57	1.80	5.02	1.87	5.64	1.41
Delayed recall rating	6.24	1.66	6.83	1.28	6.21	1.63	6.43	1.96	5.69	2.02	6.07	2.02
Serial 3s rating	6.41	2.08	5.03	2.10	5.21	2.04	4.79	2.02	4.66	1.93	4.90	2.23
Serial 7s rating	7.38	1.92	6.38	2.06	6.67	1.88	6.34	1.72	6.00	2.00	6.19	1.53
Sternberg rating	4.52	2.03	4.90	2.11	4.79	2.08	5.00	1.93	4.50	2.10	4.84	2.21
Stroop rating	4.31	1.69	4.24	1.55	4.31	1.44	4.28	1.62	4.21	1.76	4.24	1.75

3.1.5. Validity of motivation as a covariate

In order to confirm the validity of motivation as a covariate -2 Log Likelihood (-2LL) values for the LMM analysis were compared with corresponding values obtained when the same analysis was performed without the covariate. The addition of motivation improved the fit of the model as evidenced by a reduction in -2LL [57]. For comparison, LMM results including and omitting the covariate are presented in Table 2. Overall patterns of statistical significance remained largely unchanged between the two analyses.

3.2. Comparison of Cohen's d effect sizes between visits and sessions for all cognitive tasks

Cohen's d effect sizes are presented in Table 3. By comparing Cohen's d effect sizes between visits, it can be seen that practice effect sizes between Visits 2 and 3 were reduced for 9 of 11 DVs when compared with effect sizes between Visits 1 and 2. The only exceptions were DR and Stroop Interference where slight increases in practice effect sizes were observed. Practice effect sizes between same day test sessions

Table 2
LMM results for all fixed factors and covariates

LMM Fixed effects and Covariates $n = 29$

Test type	Model	Visit			Session			Visit*Session			Motivation		
		df	F statistic	p value	df	F statistic	p value	df	F statistic	p value	df	F statistic	p value
Cognitive:													
Immediate recall score (correct/15)	1	F(2,54.20)	16.82	<.001	F(1,79.39)	0.02	0.881	F(2,81.91)	4.36	0.016	F(1,136.77)	13.56	<0.001
	2	F(2,54.00)	15.91	<0.001	F(1,69.62)	1.83	0.181	F(2,82.13)	2.92	0.059	n/a	n/a	n/a
Delayed recall score (correct/15)	1	F(2,53.22)	2.79	0.070	F(1,71.43)	17.08	<0.001	F(2,65.15)	2.13	0.127	F(1,127.22)	7.64	0.007
	2	F(2,55.09)	2.87	0.065	F(1,72.68)	19.20	<0.001	F(2,67.28)	2.03	0.140	n/a	n/a	n/a
Serial 3s score (correct in 2 mins)	1	F(2,49.95)	70.48	<0.001	F(1,76.87)	42.98	<0.001	F(2,76.20)	3.87	0.025	F(1,155.27)	1.92	0.168
	2	F(5,51.08)	72.00	<0.001	F(1,74.78)	41.29	<0.001	F(2,74.32)	4.39	0.016	n/a	n/a	n/a
Serial 3s RT (ms)	1	F(2,53.89)	35.30	<0.001	F(1,86.05)	16.59	<0.001	F(2,50.31)	1.45	0.243	F(1,118.84)	2.85	0.094
	2	F(2,53.86)	33.98	<0.001	F(1,84.73)	16.13	<0.001	F(2,47.79)	1.98	0.149	n/a	n/a	n/a
Serial 7s score (correct in 2 mins)	1	F(2,46.26)	32.86	<0.001	F(1,69.02)	10.74	0.002	F(2,71.52)	0.71	0.496	F(1,135.85)	19.05	<0.001
	2	F(2,48.54)	26.88	<0.001	F(1,57.88)	9.13	0.004	F(2,67.26)	0.37	0.690	n/a	n/a	n/a
Serial 7s RT (ms)	1	F(2,66.14)	11.43	<0.001	F(1,67.33)	17.04	<0.001	F(2,73.47)	4.39	0.016	F(1,124.54)	0.05	0.825
	2	F(2,66.09)	11.37	<0.001	F(1,67.10)	17.01	<0.001	F(2,73.84)	4.39	0.016	n/a	n/a	n/a
Sternberg scanning rate (ms/item)	1	F(2,60.89)	6.87	0.002	F(1,58.29)	0.07	0.789	F(2,71.51)	0.25	0.777	F(1,130.19)	1.37	0.243
	2	F(2,60.33)	7.08	0.002	F(1,58.80)	0.10	0.754	F(2,70.08)	0.26	0.774	n/a	n/a	n/a
Sternberg RT (ms)	1	F(2,60.23)	2.82	0.067	F(1,55.60)	9.37	0.003	F(2,71.45)	1.06	0.351	F(1,128.73)	3.60	0.060
	2	F(2,60.81)	2.80	0.069	F(1,56.18)	8.80	0.004	F(2,71.12)	1.13	0.328	n/a	n/a	n/a
Stroop incongruent RT (ms)	1	F(2,52.41)	13.62	<0.001	F(1,57.19)	24.57	<0.001	F(2,74.68)	8.70	<0.001	F(1,140.55)	3.24	0.074
	2	F(2,52.04)	12.66	<0.001	F(1,58.63)	23.75	<0.001	F(2,74.99)	8.91	<0.001	n/a	n/a	n/a
Stroop congruent RT (ms)	1	F(2,43.04)	6.65	0.003	F(1,42.89)	50.08	<0.001	F(2,83.05)	3.52	0.034	F(1,140.24)	5.82	0.017
	2	F(2,48.90)	6.17	0.004	F(1,47.64)	44.81	<0.001	F(2,83.96)	3.45	0.036	n/a	n/a	n/a
Stroop interference effect (ms)	1	F(2,55.57)	0.29	0.748	F(1,61.19)	6.50	0.013	F(2,84.16)	2.40	0.097	F(1,145.5)	3.19	0.076
	2	F(2,55.68)	0.32	0.729	F(1,58.75)	5.51	0.022	F(2,84.38)	2.29	0.108	n/a	n/a	n/a
Mood:													
Mental fatigue (rating/9)	2	F(2,58.82)	5.18	0.008	F(1,69.09)	7.24	0.009	F(2,83.19)	0.52	0.599	n/a	n/a	n/a
Positive affect (score/50)	2	F(2,53.70)	5.18	0.009	F(1,61.90)	13.24	0.001	F(2,71.46)	3.13	0.050	n/a	n/a	n/a
Negative affect (score/50)	2	F(2,75.48)	0.31	0.737	F(1,58.12)	2.69	0.107	F(2,54.22)	0.40	0.673	n/a	n/a	n/a

Motivation is included as a repeated covariate in Model 1, but is omitted in Model 2.

Table 3
Cohen's d effect sizes

Cognitive Measure	Mean practice effect size between Visits			Mean practice effect size between Sessions		
				Visit1	Visit2	Visit3
	Visit1–Visit2	Visit2–Visit3	Visit1–Visit3	Session1–Session2	Session1–Session2	Session1–Session2
Immediate recall	0.368	0.333	0.686	0.395	–0.104	–0.354
Delayed recall	0.117	0.188	0.304	–0.173	–0.440	–0.650
Serial 3s score	0.351	0.258	0.608	0.434	0.311	0.113
Serial 3s RT	0.273	0.149	0.450	0.324	0.150	0.167
Serial 7s score	0.250	0.157	0.409	0.236	0.097	0.133
Serial 7s RT	0.198	0.100	0.321	0.371	0.305	0.048
Sternberg Scanning rate	0.382	0.081	0.438	0.100	–0.088	0.069
Sternberg RT	0.159	0.117	0.259	0.395	0.221	0.091
Stroop Incongruent RT	0.148	0.146	0.293	0.619	0.112	0.114
Stroop Congruent RT	0.125	0.112	0.235	0.557	0.208	0.356
Stroop Interference Effect	0.022	0.058	0.080	0.002	–0.217	–0.528

Decreases in performance are prefixed with a minus sign.

were also reduced when comparing Visits 2 and 3 with Visit 1. For IR, DR and Stroop Interference, where overall performance decreases were observed between same day sessions, the magnitude of these negative Cohen's d values increased at subsequent visits.

3.3. LMM analysis of the changes in mood performance across all test sessions

Mean mood scores with standard deviations recorded at each time point are presented in Table 1. The LMM model results are summarised in Table 2. For clarity, only significant fixed effects are reported in the text.

Mental fatigue was predicted by Visit $[F(2,58.82) = 5.18, \ p = 0.008]$ with *post hoc* comparisons revealing decreased ratings of fatigue at Visit 3 compared with both Visit 1 $[p = 0.077]$ and Visit 2 $[p = 0.008]$. Session was also a significant factor $[F(1,69.09) = 7.24, \ p = 0.009]$, with increased mental fatigue observed at Session 2 relative to Session 1.

Positive affect was significantly predicted by Visit $[F(2,53.70) = 5.18, p = 0.009]$, with *post hoc* comparisons revealing decreased mood at Visit 2 compared with Visit 1 $[p = 0.006]$. Significant decreases in positive affect across same day test sessions were indicated by the factor Session $[F(1,61.90) = 13.24, p = 0.001]$. Moreover, the interaction between Visit and Session was also significant $[F(2,71.46) = 3.13, p = 0.050]$. Significant decreases in positive affect between sessions were found at Visit 1 $[p = 0.001]$ and Visit 2 $[p < 0.001]$ only following *post hoc* comparisons.

Reassuringly, no significant changes in negative affect were observed throughout testing.

3.4. LMM analysis of the changes in perception of task difficulty for all cognitive tasks, across all test sessions

Figure 2 represents mean difficulty ratings for all cognitive tasks at each testing time point.

Cognitive task type was a significant factor in predicting difficulty ratings $[F(5,249.55) = 60.84, p < 0.001]$. In task order, Serial 7s was rated the most difficult, followed by DR, IR, Serial 3s, Sternberg and Stroop. However not all differences were statistically significant. *Post hoc* comparisons in order of difficulty revealed trends or significant differences between DR and IR $[p = 0.002]$, IR and Serial 3s $[p = 0.087]$, and Sternberg and Stroop $[p = 0.055]$.

4. Discussion

The aim of this study was to investigate the effects of repeated exposure to a task on cognitive performance, within and between test days. In summary, practice effects remained evident either across weekly visits or between same-day sessions for all of the cognitive tasks investigated despite the use of alternate forms of each task and familiarisation trials immediately before the start of each test performance. However, in accordance with previous findings [10, 13, 16] the strongest practice effects appeared between visits 1 and 2. So a separate familiarisation visit, followed by a period of consolidation before data collection, is likely to aid in addressing
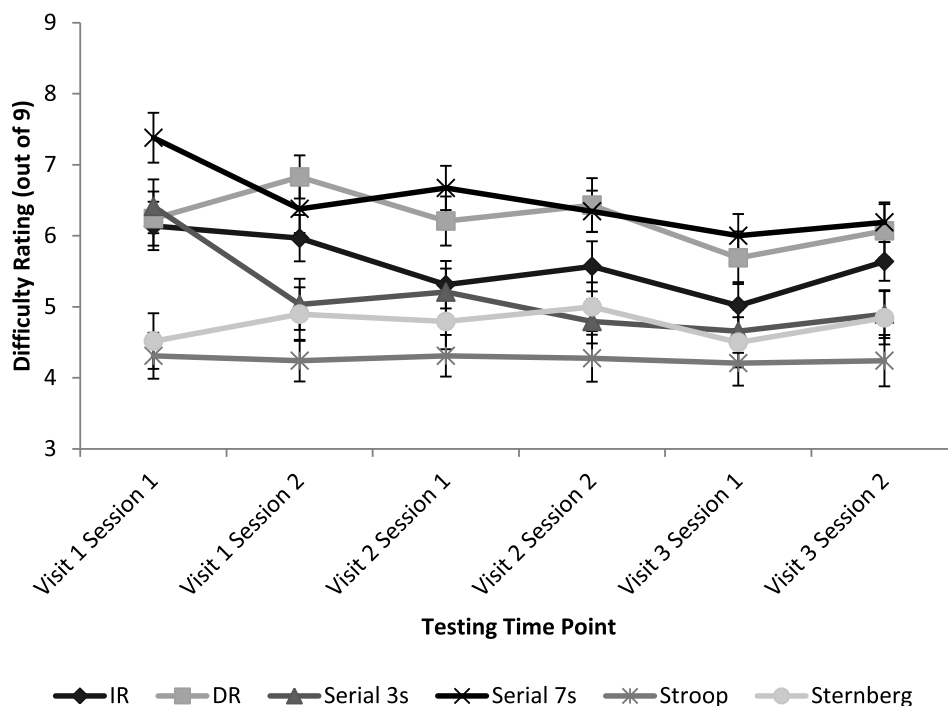
Fig. 2. Mean subjective ratings of task difficulty for each cognitive task at each testing time point. Error bars represent mean standard error.

the impact of practice related improvements on test performance between visits. Specifically, significant practice effects were not evident between visits 2 and 3 for the Sternberg task and, although still significant, were attenuated for IR, serial 3s, serial 7s and Stroop when compared with performance increases observed between visits 1 and 2.

Practice effect sizes between same day sessions were similarly reduced at visits 2 and 3 compared with visit 1. However the results suggest that familiarisation trials, whether on a separate visit or immediately prior to data collection, are likely to be ineffective at actually eliminating practice effects between same day test sessions. A similar observation was made by Lemay et al. [33]. Indeed, the lack of a *Visit*Session* interaction for the majority of the DVs reported here suggests that performance increases across multiple testing points on a single day remain relatively constant between visits, with only slight attenuation of effect sizes at subsequent visits. Residual between-session practice effects were evident only as small Cohen's d effect sizes ($d = 0.05$–0.35). Nevertheless in nutrition intervention studies, nutrient effect sizes typically range from small to moderate. For example, effect sizes following acute flavonoid intervention have been reported to range upwards from $d = 0.16$ [1]. There-

fore even small practice effect sizes may impact statistical power in a nutrition intervention study and should be taken into account in a priori power calculations.

The statistically significant improvements in task performance between test days shown here remained apparent when the variance accounted for by motivation was included in the model. By including motivation, this supports the notion that the observed effects are likely related to practice, although other variables such as mood or fatigue cannot be ruled out. Mood and fatigue were not added to the model here as measurements were recorded once per session of cognitive testing and not following individual cognitive tasks, however future studies should seek to investigate their potential impact. As a covariate, motivation was found to significantly predict episodic memory (IR and DR) recall, serial 7s and Stroop task performance. With the exception of Stroop, these tasks were also rated to be the most difficult tasks overall. Therefore, motivation was a strong predictor of performance when tasks were particularly difficult, independently of practice. Motivation was reported retrospectively and so it is possible that participants reported high motivation when they felt they had performed well during the task, even though direct feedback was not given. However if this was the case,

motivation might be expected to have been a significant predictor of performance for all of the cognitive tasks investigated. Fatigue is also likely to be a major contributing factor to task performance. As expected following an extended period of cognitive testing, mental fatigue was observed to increase at session 2 relative to session 1 and may also have impacted subjective ratings of motivation. Perceptions of task difficulty were observed to decrease over time for all cognitive tasks, suggestive of a general practice effect. After repetition, the tasks were rated as becoming easier and performance on the tasks improved. Minor fluctuations in this downward trend in perceived difficulty are again likely to result from the impact of fatigue on participants' retrospective, subjective ratings.

Depending on the cognitive task, different performance related effects of repeated exposure were observed. It has been demonstrated that practice related effects are evident across separate visits and same day sessions of testing, particularly for tasks measuring working memory (serial subtraction). For executive function tasks such as Stroop interference effect and episodic memory tasks such as immediate and delayed recall, slight practice-related improvements were observed across visits, particularly between visits 1 and 2, but apparent decreases in performance were observed between same day test sessions. For the IR and DR episodic memory tasks there is some suggestion of interference at the second session of each visit, that counters practice related improvements. The LMM analysis revealed that motivation significantly predicted memory performance and so motivational decreases are likely to account for some of the observed decreases in performance between these same day sessions, although as discussed above, motivation may be confounded with mental fatigue. Word list interference from the previous session is also likely to be a major contributing factor here [58, 59].

Performance decreases on the Stroop interference measure are more complex. In fact, overall reaction times decreased between sessions, but more so for the congruent trials, resulting in an increased interference effect. This is a potentially interesting observation that warrants further investigation, particularly as previous research has often, though not always, shown the Stroop interference effect to be reduced by practice [60]. Therefore, it is important to interpret Stroop interference effect changes with reference to the underlying reaction times. Interestingly, for all of these tasks where performance

decreases were observed between same day sessions, Cohen's d effect sizes increased at all subsequent visits. This appears consistent with the notion that practice related improvements had been attenuated and therefore interfering effects now demonstrated greater magnitude in the opposing direction.

The Sternberg task appears relatively robust to practice compared to the other cognitive tasks investigated here; an observation shared by previous research [45–47]. Overall RT performance did speed up slightly due to practice, which may be explained by faster visual processing, decision making or motor output. However the main DV for the task measures how fast participants can scan items in their STM, and the above cited research shows this to be a fixed time per item that cannot be changed by practice, reflecting a fundamental cognitive process. Despite this, the observational data here suggest some practice related improvement between visits 1 and 2 but performance appeared to stabilise after this. For the subjectively reported mood measures, as might reasonably be expected after an extended period of cognitive testing, higher ratings of mental fatigue were observed at Session 2 relative to Session 1 within visits. Similarly, a decrease in positive mood was observed at Session 2. However, no corresponding increases in negative affect were observed. These measures imply no serious negative consequences for participants in terms of fatigue, or mood over prolonged periods of repeated testing.

There are some shortcomings to this study. The effects of practice have not been investigated beyond a period of three weeks; the findings are exclusive to the tests examined; and the effects of including dual baselines [61], such as those used by Scholey et al. [52], have not been investigated. Dual baseline designs discard the first session of data collected on each visit. However, the findings of the current study demonstrate that practice cannot be assumed to have been adequately addressed through the use of alternate forms of stimuli or familiarisation trials immediately prior to data collection, and suggest that dual baseline designs may be similarly limited in their ability to eliminate practice effects.

Contrary to previous findings [10, 16] but consistent with theories of learning and memory, within a nutrition intervention design framework, the effects of practice have been observed to persist well past the first two repeat sessions of cognitive testing on the majority of tasks investigated here. Although the Sternberg memory scanning task appeared to be an exception, becoming relatively resistant to

practice-related improvements after an initial test visit. Practice can interact with a nutrition intervention such that more learning takes place in the active condition and then what is learned is retained and causes elevated performance in the control condition, masking intervention effects for half the participants in a typical counterbalanced crossover design. Conversely, care must be taken to ensure that improvements in performance arising from repeated practice of the task when the control condition is performed first are not confused with those associated with the intervention. Therefore the use of practice resistant tasks or appropriate methodology, as outlined below, for minimising the influence of practice effects is paramount. Motivation and fatigue clearly also contribute to cognitive performance, particularly for more difficult cognitive tasks. The relationship between motivation, fatigue and practice requires further investigation.

The findings of this study are likely to generalise as similar memory, reaction time and accuracy components are inherent in all commonly used cognitive tasks. In future intervention studies, whether that be nutritional or pharmacological, it is recommended that methodologies take practice effects into account; a summary of recommendations is included in the final paragraph. Additionally, it is recommended that researchers and cognitive task developers continue to collect normative data related to repeat testing, in order to assess normal practice-related learning curves across a range of cognitive domains and testing intervals. For studies which have already reported effects of nutrition interventions, care should be taken when retrospectively calculating effect sizes. For example in the absence of an appropriate control condition, effect size may be inflated by practice. On the other hand, a study may be underpowered as a result of the additional variance introduced by practice [21]. In the absence of good design and/or replication, effect sizes for nutrition interventions should be accepted with caution [1].

Methodological recommendations emerging from the study include the importance of an initial familiarisation visit which is on a separate day to subsequent testing. Indeed, further investigation is required to determine the impact of the length of time between familiarisation and test days. The inclusion of a baseline at each visit within a counterbalanced framework including the control, and the use of tasks with a reduced susceptibility to practice such as Sternberg is also recommended. Alternate forms and familiarisation trials are also recommended at

each task performance, along with a between participants counterbalanced order of tasks within a battery. As these measures will only minimise, not eliminate practice effects, residual practice effects should always be taken into account in the design stage when selecting the number of participants required to adequately power a crossover intervention study. Additionally practice effects may be taken into account in the data analysis stage. For example, using LMM analysis it is possible to include testing order as a factor within the analysis model. Parallel designs, with a separate control group of participants, may be used to mitigate practice effects; or even exploit them, as in ageing research where subjects with cognitive deficits may not show typical practice effects [62]. Generally however, such designs will add between-subjects error variance and increase the number of participants needed. An active intervention may also interact with practice, for example through improved strategy selection. Therefore it is still recommended to address practice effects in a parallel design if there is a repeated testing component, such as with longitudinal designs. In combination, these measures will aid in improving the reliability and validity of repeated cognitive testing in nutritional intervention designs.

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Bell L, Lamport DJ, Butler LT, Williams CM. A review of the cognitive effects observed in humans following acute supplementation with flavonoids, and their associated mechanisms of action. Nutrients. 2015;7:10290-306.

[2] Blumberg JB, Ding EL, Dixon R, Pasinetti GM, Villarreal F. The Science of Cocoa Flavanols: Bioavailability, Emerging Evidence, and Proposed Mechanisms. Adv Nutr. 2014;5: 547-9.

[3] Lamport DJ, Saunders C, Butler LT, Spencer JP. Fruits, vegetables, 100% juices, and cognitive function. Nutr Rev. 2014;72:774-89.

[4] Mecocci P, Tinarelli C, Schulz RJ, Polidori MC. Nutraceuticals in cognitive impairment and Alzheimer's disease. Front Pharmacol. 2014;5:1-11.

[5] Nehlig A. The neuroprotective effects of cocoa flavanol and its influence on cognitive performance. Br J Clin Pharmacol. 2013;75:716-27.

[6] Kumar H, More SV, Han SD, Choi JY, Choi DK. Promising therapeutics with natural bioactive compounds for improving learning and memory - A review of randomized trials. Molecules. 2012;17:10503-39.

[7] Shukitt-Hale B. Blueberries and neuronal aging. Gerontology. 2012;58:518-23.

[8] Williams RJ, Spencer JPE. Flavonoids, cognition, and dementia: Actions, mechanisms, and potential therapeutic utility for Alzheimer disease. Free Radic Biol Med. 2012;52:35-45.

[9] Gupta C, Prakash D. Nutraceuticals for geriatrics. J Tradit Complement Med. 2015;5:5-14.

[10] Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. J Clin Exp Neuropsychol. 2006;28:1095-112. doi: 10.1080/13803390500205718

[11] Hausknecht JP, Halpert JA, Paolo NT Di, Gerrard MOM. Retesting in Selection: A Meta-Analysis of Practice Effects for Tests of Cognitive Ability. J Appl Psychol. 2007;92: 373-85.

[12] Lamport DJ, Dye L, Wightman JD, Lawton CL. The effects of flavonoid and other polyphenol consumption on cognitive performance: A systematic research review of human experimental and epidemiological studies. Nutr Ageing. 2012;1:5-25.

[13] Collie A, Maruff P, Darby DG, McStephen M. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. J Int Neuropsychol Soc. 2003;9:419-28.

[14] McCaffrey RJ, Ortega A, Haase RF. Effects of repeated neuropsychological assessments. Arch Clin Neuropsychol. 1993;8:519-24.

[15] Basso MR, Bornstein RA, Lang JM. Practice effects on commonly used measures of executive function across twelve months. Clin Neuropsychol. 1999;13:283-92.

[16] Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. BMC Neurosci. 2010;11:118. doi: 10.1186/1471-2202-11-118

[17] McClelland GR. The effects of practice on measures of performance. Hum Psychopharmacol Clin Exp. 1987;2:109-18.

[18] Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. Neuropsychologia. 1998;36:915-23.

[19] Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD. Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: Variability and change with practice. J Neurosci. 2003;23:5945-52.

[20] Petersen SE, van Mier H, Fiez JA, Raichle ME. The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci. 1998;95:853-60.

[21] McCaffrey RJ. Serial neuropsychological assessment with the National Institute of Mental Health (NIMH) AIDS Abbreviated Neuropsychological Battery. Arch Clin Neuropsychol. 2001;16:9-18.

[22] EFSA. Guidance on the scientific requirements for health claims related to functions of the nervous system, including psychological functions. EFSA J. 2012;10:1-13. doi: 10.2903/j.efsa.2012.2816

[23] de Jager CA, Dye L, de Bruin EA, Butler L, Fletcher J, Lamport DJ, Latulippe ME, Spencer JPE, Wesnes K. Criteria for validation and selection of cognitive tests for investigating the effects of foods and nutrients. Nutr Rev. 2014;72:162-79.

[24] Beglinger L, Gaydos B, Tangphao-Daniels O, Duff K, Kareken D, Crawford J, Fastenau P, Siemers E. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol. 2005;20:517-29. doi: 10.1016/j.acn.2004.12.003

[25] Roebuck-Spencer T, Sun W, Cernich AN, Farmer K, Bleiberg J. Assessing change with the Automated Neuropsychological Assessment Metrics (ANAM): Issues and challenges. Arch Clin Neuropsychol. 2007;22:S79-S87.

[26] Kennedy DO, Jackson PA, Haskell CF, Scholey AB. Modulation of cognitive performance following single doses of 120mg Ginkgo biloba extract administered to healthy young volunteers. Hum Psychopharmacol. 2007;22:559-66.

[27] Alharbi MH, Lamport DJ, Dodd GF, Saunders C, Harkness L, Butler LT, Spencer JPE. Flavonoid-rich orange juice is associated with acute improvements in cognitive function in healthy middle-aged males. Eur J Nutr. 2016;55:2021-29.

[28] Cox KH, Pipingas A, Scholey AB. Investigation of the effects of solid lipid curcumin on cognition and mood in a healthy older population. J Psychopharmacol. 2014; doi: 10.1177/0269881114552744

[29] Kennedy DO, Scholey AB, Wesnes KA. Modulation of cognition and mood following administration of single doses of Ginkgo biloba, ginseng, and a ginkgo/ginseng combination to healthy young adults. Physiol Behav. 2002;75:739-51.

[30] Wightman EL, Haskell CF, Forster JS, Veasey RC, Kennedy DO. Epigallocatechin gallate, cerebral blood flow parameters, cognitive performance and mood in healthy humans: A double-blind, placebo-controlled, crossover investigation. Hum Psychopharmacol. 2012;27:177-86.

[31] Rigney U, Kimber S, Hindmarch I. The effects of acute doses of standardized Ginkgo biloba extract on memory and psychomotor performance in volunteers. Phytother Res. 1999;13:408-15.

[32] Wesnes K, Pincock C. Practice effects on cognitive tasks: A major problem? Lancet Neurol. 2002;1:473. doi: 10.1016/S1474-4422(02)00236-3

[33] Lemay S, Bedard M-A, Rouleau I, Tremblay PG. Practice Effect and Test-Retest Reliability of Attentional and Executive Tests in Middle-Aged to Elderly Subjects Practice Effect and Test-Retest Reliability of Attentional and Executive Tests in Middle-Aged to Elderly Subjects. Clin Neuropsychol. 2004;18:1-19.

[34] Alharbi MH, Lamport DJ, Dodd GF, Saunders C, Harkness L, Butler LT, Spencer JPE. Flavonoid rich orange juice is associated with acute improvements in cognitive function in healthy middle-aged males. Eur J Neurosci. 2015;1-9. doi: 10.1007/s00394-015-1016-9

[35] Nathan PJ, Ricketts E, Wesnes K, Mrazek L, Greville W, Stough C. The acute nootropic effects of Ginkgo biloba in healthy older human subjects: A preliminary investigation. Hum Psychopharmacol. 2002;17:45-9.

[36] Watson AW, Haskell-Ramsay CF, Kennedy DO, Cooney JM, Trower T, Scheepens A. Acute supplementation with black-

currant extracts modulates cognitive functioning and inhibits monoamine oxidase-B in healthy young adults. J Funct Foods. 2015;17:524-39.

[37] Dikmen SS, Heaton RK, Grant I, Temkin NR. Test–retest reliability and practice effects of Expanded Halstead–Reitan Neuropsychological Test Battery. J Int Neurpsychological Soc. 1999;346-56.

[38] Anderson JR. Automaticity and the ACT-super(*) theory. Am J Psychol. 1992;105:165-80.

[39] Macready AL, Butler LT, Kennedy OB, Ellis JA, Williams CM, Spencer JPE. Cognitive tests used in chronic adult human randomised controlled trial micronutrient and phytochemical intervention studies. Nutr Res Rev. 2010;23:200-29.

[40] Macready AL, Kennedy OB, Ellis JA, Williams CM, Spencer JPE, Butler LT. Flavonoids and cognitive function: A review of human randomized controlled trial studies and recommendations for future studies. Genes Nutr. 2009;4:227-42.

[41] Whyte AR, Schafer G, Williams CM. Cognitive effects following acute wild blueberry supplementation on 7- to 10-year-old children. Eur J Nutr. 2015;1-12. doi: 10.1007/s00394-015-1029-4

[42] Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol. 1935;18:643-62.

[43] Sternberg S. Memory-scanning - Mental processes revealead by reaction-time experiments.pdf. Am Sci. 1969;57:421-57.

[44] Sternberg S. High-Speed Scanning in Human Memory. Sci New Ser. 1966;153:652-4.

[45] Sternberg S. Memory scanning: New findings and current controversies. Quartely J Exp Psychol. 1975;27:1-32.

[46] Kristofferson MW. The effects of practice with one positive set in a memory scanning task can be completely transferred to a different positive set. Mem Cognit. 1977;5:177-86. doi: 10.3758/BF03197360

[47] Kristofferson MW. Effects of Practice on Character-Classification Performance. Can J Psychol. 1972;26:54-60.

[48] Böcker KBE, Hunault CC, Gerritsen J, Kruidenier M, Mensinga TT, Kenemans JL. Cannabinoid modulations of resting state EEG θ power and working memory are correlated in humans. J Cogn Neurosci. 2010;22:1906-16.

[49] Grattan-Miscio KE, Vogel-Sprott M. Effects of alcohol and performance incentives on immediate working memory. Psychopharmacology (Berl). 2005;181:188-96.

[50] Vinkhuyzen AAE, van der Sluis S, Boomsma DI, de Geus EJC, Posthuma D. Individual differences in processing speed and working memory speed as assessed with the Sternberg memory scanning task. Behav Genet. 2010;40:315-26.

[51] Subhan Z, Hindmarch I. The psychopharmacological effects of Ginkgo biloba extract in normal healthy volunteers. Int J Clin Pharmacol Res. 1984;4:89-93.

[52] Scholey AB, French SJ, Morris PJ, Kennedy DO, Milne AL, Haskell CF. Consumption of cocoa flavanols results in acute improvements in mood and cognitive performance during sustained mental effort. J Psychopharmacol. 2010;24: 1505-14.

[53] Scholey AB, Harper S, Kennedy DO. Cognitive demand and blood glucose. Physiol Behav. 2001;73:585-92.

[54] Scholey AB, Bauer I, Neale C, Savage K, Camfield D, White D, Maggini S, Pipingas A, Stough C, Hughes M. Acute effects of different multivitamin mineral preparations with and without Guaraná on mood, cognitive performance and functional brain activation. Nutrients. 2013;5:3589-604.

[55] Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Pers Soc Psychol. 1988;54:1063-70.

[56] Corbin L, Marquer J. Individual differences in Sternberg's memory scanning task. Acta Psychol (Amst). 2009;131: 153-62.

[57] Shek DTL, Ma CMS. Longitudinal Data Analyses Using Linear Mixed Models in SPSS: Concepts, Procedures and Illustrations. Sci World J. 2011;11:42-76.

[58] Greenberg R, Underwood BJ. Retention as a function of stage of practice. J Exp Psychol. 1950;40:452-7.

[59] Underwood BJ. Interference and forgetting. Psychol Rev. 1957;64:49-60.

[60] Macleod CM. Half a Century of Research on the Stroop Effect: An Integrative Review. Psychol Bull. 1991;109:163-203.

[61] McCaffrey RJ, Ortega A, Orsillo SM, Nelles WB, Haase RE. Practice effects in repeated neuropsychological assessments. Clin Neuropsychol. 1992;6:32-42.

[62] Krenk L, Rasmussen LS, Siersma VD, Kehlet H. Short-term practice effects and variability in cognitive testing in a healthy elderly population. Exp Gerontol. 2012;47:432-6.