RESEARCH ARTICLE

Effects of practice on individual difference in tasks employed to assess cognitive abilities

Daniel Gopher[1], Daniel Ben-Eliezer[2], Paul D. Feigin[3]

[1]Professor Emeritus of cognitive psychology and human factors engineering; Faculty of Data and Decision Sciences; Technion-Israel Institute of Technology. Haifa
[2]Post-doctoral student in cognitive psychology; Faculty of Data and Decision Sciences.; Technion Faculty of Industrial Engineering 2018 (cognitive psychology, human factors)
[3]Emeritus Professor of Statistic; Faculty of Data and Decision Sciences; Australian National University: PhD 1975

OPEN ACCESS

ABSTRACT

The study examines the influence of the format and features e of tasks employed to assess cognitive abilities. Three experiments investigated the effect of practice on performance differences between performers of these tasks. Two experiments were conducted on a computerized, demanding task, developed to assess attention management and multitask performance. The third experiment examined practice effects in six replications of a five tasks battery, consisting of tasks commonly used in the evaluation of cognitive functions. Significant individual differences were observed in all experiments, within and across tasks. However, practicing differentially affects tasks and components within tasks.

Three types of practice effects on performance were identified: (1) Performance levels did not change and did not benefit from practice in components which mainly draw upon the operation of bottom-up, exogenous attention systems; (2) Practice had strong asymmetric effects on performance when performance required to resolve a conflict between two automatically attended elements. In these tasks, performers with lower initial performance scores benefited more from practice than those with higher initial performance levels. (3) For tasks in which executive control and working memory were called upon but there was no conflict to resolve, significant practice effects were obtained with equal gains for performers differing in initial performance levels.

The important implication of the obtained results is that when a task is developed as test to evaluate a cognitive ability, its format and features may affect first session performance and only clarified with practice. The contribution of format and components are discussed with reference to research on the evaluation of individual differences in cognitive abilities, types of attention demand and working memory requirements.

Keywords: Practice effects, cognitive tasks, Individual differences, attention systems

Introduction

Individual differences in cognitive abilities are evaluated across a wide range of applications and problem areas. Examples are: admission to education and training programs (Oren et al., 2014[1]; Kleper, & Seka, 2017[2]); job selection (Gopher, 1982[3]; Charles & Florah, 2021[4]); developmental progress assessments (Friedman et al., 2016[5]; Karbach et al., 2017[6]); aging (Salthouse & Ferrer-Caja, 2003[7]; Eich et al., 2016[8]); brain dysfunctions (Bialystok et al., 2008[9]; Tanguay et al., 2014[10]); rehabilitation (Folstein et al., 1975[11]; Lucas et al., 1998[12]). The increasing interest and focus have been accompanied by the introduction of a growing number of commercial cognitive test batteries, which are employed to assess individual differences in cognitive abilities and functions (e.g., MoCA, CANS-MCI, MMSE, DRS-Mattis, USC-ADRC). What is common to these batteries are the conduct of a single or two test sessions, within which participants perform once or twice an arsenal of specified cognitive tasks. Scores of each participant are compared to others (e.g., in one's group, designated populations, norms, or criteria). Cutoff points are set and decisions drawn based upon these scores. However, it is well recognized that although statistically significant, the correlations between test and criteria usually fall in the .3 to .7 range. thereby reducing the effectiveness of the test as a sole measure of the targeted cognitive ability (Salthouse, 2005[13]; Rose et al, 2015[14]; Kabat et al, 2001[16]). One possible reason for lower correlation is that there may be other contributing factors in addition to the ability measured by the test. A second possible contributing factor stems from the fact that in addition to its prime intended objective, each test is a task enveloped and influenced by its specific format, sensory features, perceptual and response modes, as well as instructions and performance procedures. Researchers of individual differences in cognitive abilities have been aware of this fact. In the theoretical and conceptual frameworks which have been developed to specify types and dimensions of cognitive ability, single tasks have been grouped and associated through their loading on underlying factors, dimensions or cognitive structure. Each of the tasks has been shown to be mainly loaded on one underlying cognitive dimension but may also correlate with other dimensions (Miyake et al., 2000[17]; Salthouse, 2005[13]; Ackerman & Cianciolo, 2000[18]).

One common approach to group tasks is factor analysis. A good example is the multiple studies conducted by Salthouse and his colleagues (Salthouse, 2004[19], 2005[13]). They proposed an analytical model with five cognitive functions: vocabulary, reasoning, space, memory and speed. Each was mapped to a group of 3 – 4 tests, from sixteen individual cognitive tasks (e.g., Reasoning, Ravens, Shipley Abstractions, Letter Set). Correlational and factor analysis computed the loading of each test on its mapped cognitive dimension. Dimension levels were then compared along age and gender. A later study (Salthouse, 2005[13] applied the same format and added tasks of executive control (task switching, inhibition and working memory).

A related second approach is latent variable analysis, developed and investigated in multiple studies by Miyake, Friedman and their collaborators (Miyake et al., 2000[17]; Miyake, & Friedman, 2012[20]; Friedman, & Miyake, 2017[21]). The focus of this group has been on executive control processes. Accordingly, they proposed three control functions: updating, shifting and inhibition. Each mapped into a group of tasks given to subjects (e.g., shifting between colors, numbers, category). They employ confirmatory factor analysis (CFA) and structural equation modeling (SEM) to compute the correlation between tasks and their respective latent dimension, as well as the interrelations between the three control functions, representing the "unity" of executive control.

The present study examines whether lack of practice could be a contributor to the lower correlation of a single task performance measure with its targeted intended cognitive ability. As indicated, in cognitive test batteries, each test is commonly presented once or twice, but there is no practice or training. No or limited practice is also common to experimental

studies developed to distinguish between groups in a cognitive ability measure. However, when first presented with the targeted task, task performance may reflect the targeted ability but also the experience with one or many of the task format and features, which may be relevant or irrelevant to the measurement of the targeted ability (e.g. working memory).

How will the difference between performers on task performance change with practice? Practice is in providing a common experience on the task to reduce the effects of initial variability due to previous experience with its features and format. If performers' scores improve above a set criterion level, or performers change their relative position up or down in the tested group, what will be a more valid estimate of the cognitive function targeted by the task? Overall, practice may have four possible effects on performance: (1) If there is no practice effect, then task performance before and after practice equally reflects individual differences in the task. (2) Practice improves the performance of all individuals and there is equal gain for first session high and low performers. In this case, the absolute performance levels following practice are important. For example, in the evaluation of cognitive impairment level, developmental progress and meeting selection criteria, both final performance levels and rate of improvement are informative. Types 3 and 4 of effects, consider possible interaction between evaluation of individual differences in the first and last administration of the test. (3) If lower performers in the first session gain more from practice than higher initial performers, they improve their performance and position in the tested group relative to their rank in the first session. Thus, when assessing individual differences, practice may change both performance levels and relative rankings of low compared to high initial performers. (4) The opposite case is where practice causes higher initial performers to gain more than low initial performers. In this case, practice does not change the relative position of individual subjects but increases the spread of the initial performance differences between high and low performers observed during the first administration of the task.

Taken together, in all four cases practice may inform us of the nature of the task constructed and its employment as an estimate of its targeted cognitive ability.

Differential influence of practice on task performance has been reported in training research which investigated the influence of training manipulations on different ability groups (Frederiksen & White, 1989[22]; Gopher, Weil & Siegel 1989[23]). A more recent line of experiments studied the relationship between task switching training and individual differences by comparing before and after practice cognitive ability tests, reporting asymmetric effect of training on individual differences in cognitive tasks, (Karbach et al., 2017[6]; Zinke et al., 2012[24]; Zinke et al., 2014[25]; Karbach & Kray, 2009[26]; Bherer et al., 2008[27]; Cepeda et al., 2001[28]). They showed that individuals with lower cognitive abilities at pretest showed larger training and transfer benefits after the training. They labelled it "compensation" effect, in which baseline cognitive abilities were negatively correlated with the training-induced gain (i.e., low-performing individuals benefitted more). Similar compensation effects were reported for working memory, executive control and a variety of cognitive tasks. In all cases training gains were compared with a "no training" control group (Zinke et al., 2012[24], 2014[25]; Karbach et al., 2009[26]; Beherer et al., 2008)[27]. Karbach et al. (2017)[6] tested three age groups (children, young and older adults). Zinke et al (2012)[24] tested older adults (age 77 – 96), and in Zinke et al (2014)[25] the sample ages were between 65 to 95 years. Despite significant differences in average performance between age groups, the compensation effect was found in all of them, thus eliminating ceiling or regression to the mean explanations of the training effect (as also described in: Konen & Karbach, 2021[29]).

Opposing the compensation influence of training, there is also a line of studies reporting positive magnification effects. That is, subjects with high initial ability, benefit more from training than low ability

subjects, leading to positive rather than negative correlation between intercept and training slope (Baltes & Kliegl, 1992[30]; Brehmer et al, 2007[31]; Foster et al., 2017[32]; Lindenberger et al., 1992[33]; Lövdén et al., 2012[34]; Verhaeghen & Marcoen, 1996[35]). These studies investigated the training of high and low working memory subjects when teaching them strategies and techniques for improving working memory span (e.g., mnemonics). Positive correlations imply that in terms of cognitive testing, the scores and individual differences obtained at the first presentation of test, are a good representation of the differences between subjects on the targeted cognitive ability (e.g., working memory span, processing speed). In this case, training would clarify and increase the range and the distinction between individuals but would not change their relative position. However, if test performance is used with reference to a cutoff point, diagnosis or graduation criterion, training may change the status of individuals. In this regard, the question again is what the added value of the post-training is compared with the pre-training score.

It is important to emphasize at this point the difference between training and practice. Training studies investigate the influence of training manipulation or intervention on performers differing in their initial task performance levels. In practice only there is no manipulation, performers accumulate repeated spontaneous experience on a task format and features as is. Practice performance slope and the first and last performance levels are compared. Training applies to the influence of teaching or instruction method, which most often is compared with control groups and with transfer to external tasks. Hence, compensatory and magnification training studies are supportive but different from the investigation of the effects of practice on targeted tests performance levels.

We report the results of three experiments, which investigated the influence of practice on performance and differences between performers. Two experiments were conducted on two versions of the computerized Breakfast Task, developed by Craik and Bialystok (2006)[36] to test attention management and multitask performance in a simulated daily mission. Subjects in both experiments were given five sessions of practice. The third experiment examined six replications of a cognitive test battery comprising five commonly used cognitive tests.

2. Tasks and Methods

2.1 BREAKFAST TASK (ALTERNATING SCREENS)
It is a computer-based simulation, in which the performer is required to cook five food items for breakfast, while concurrently setting as many tables as possible for four guests (Figure 1). In its original form, it was developed by Craik and Bialystok (2006)[36] as an indicator of coping with multitasking. high executive control and attention management demands. It also aspired to simulate a daily activity.

They employed the task to compare planning and attention management capabilities of young and older adults and contrasted monolingual and bilingual performers in each age group. Older bilinguals showed advantage over older monolinguals. In later studies, the task was applied to compare young, older and Parkinson's disease patients (Bialystok et al., 2008)[9] and to evaluate the impact of brain injury (Tanguay et al., 2014)[10]. In all their experiments they noted the performance difference between cooking and table setting, but their main interest was in the comparisons between specified groups. The task was conceived as an integrated index of multitasking ability and only limited, or no practice was given. All studies compared group averages, with considerable variability in each average, indicating individual differences. Task difficulty was increased, but there was no administrated practice.

2.1.1 Participants: Thirty participants were recruited to a five-session practice program. All participants were students of the Technion, aged between 19 and 29 (mean age: 22.8). 60% of them were females and had normal or corrected vision. After finishing

the task, participants received 250 NIS (about 70 USD). Three participants who received the highest combined standardized scores in a session received 50 NIS bonuses for the five sessions. Participants were recruited to the experiment by responding to a post addressed to students of the Technion which included an online questionnaire. Candidates with excessive experience of playing daily video or computer games for three hours or more and those with very limited computer use were excluded. All participants gave written informed consent according to the declaration of Helsinki guidelines, as approved by the Technion ethics committee.
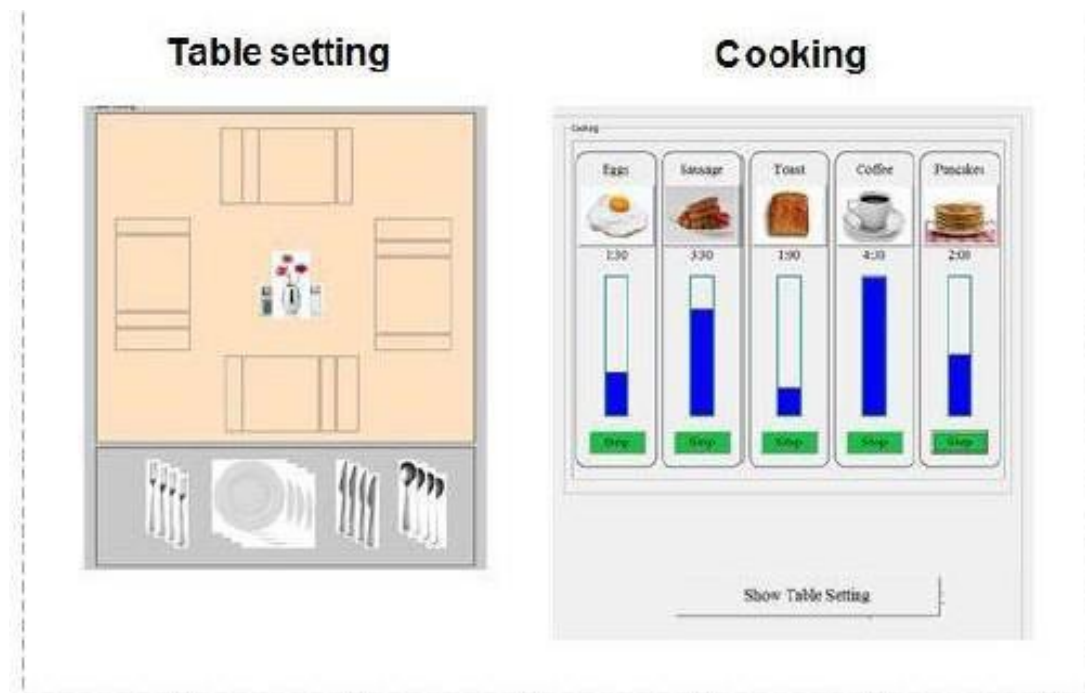
2.1.2 Design: Participants were required to cook food for breakfast while concurrently setting tables for four guests (Figure 1). In the cooking segment, participants were asked to cook five food items; cooking time in minutes for each item was displayed (coffee – 4.5, sausage – 3.5, pancakes – 2, egg – 1.5, and toast – 1). Each cooking round started with the longest duration item (coffee) and additional food items were added from among the other four. Ideally, food items should be accurately cooked and all items completed together, cooking times should be ordered from long to short items and the starting time of each should be calibrated relative to the starting time of the previous food item. The barographs in Figure 1 indicate the optimal starting time for each food item to be accurately cooked and served together.

Three performance aspects were measured. Cooking Time Discrepancy *(DIS):* the average absolute values of the differences between the required and actual cooking times of each item. Range of stop times (RST); the difference in seconds between the first and last food item to complete cooking; For table setting (TS), participants had to set tables for four guests, which included placing the plate, fork, knife and spoon in their locations. Table setting could follow two placement rules: By guest – the complete set should be placed for a guest, before proceeding to the next guest; or by utensil – each type of utensil should be placed for all four guests

(i.e., four forks, then four knives…). The program did not allow putting a utensil in the wrong place. When a table setting is completed for four guests, a new table is presented. Table setting scores counted only fully set tables based on placement instructions. Feedback on DIS, RST and TS were displayed at the end of each cooking round.

2.1.3 Procedure: Following signing an Informed Consent form, each participant received a personal username allowing them to log in to the Breakfast Task server online and perform the task from their home. In the first login, they filled in a demographic inventory, read instructions, and performed two demonstration trials. They were then instructed to finish their five practice sessions within 14 days and were prevented from performing two sessions in one day. The research assistant was in continuous touch with participants via phone, monitoring their progress and providing technical assistance.

Practice included five sessions, each with eight cooking rounds each lasting 4.5 minutes (40 rounds total) and with identical structure: Cooking five food items, presented on the same screen and setting tables for four guests, with alternating setting rule – three tables by guest followed by three tables by utensil and so on; the table setting and cooking display screens were alternating (Craik & Bialystok, 2006[36], condition 2), participants had to click on a button, to switch between the table and cooking screens. Cooking and table setting performance were instructed to be of equal priority and task performance is evaluated by the combined performance on both.

Figure 1: *The two Segments of the Breakfast Task.*

Note. Bars display the start and elapsed cooking time of each cooking item. The bars in the figure display the optimal starting time for each item. Cooking time for each item is displayed under its image.

2.2 BREAKFAST TASK (SIX SCREENS)

2.2.1 Task and design: Five sessions of the Breakfast Task were administer red identically as in the first Experiment, with one major difference: alternating between tables display but only one of five food items, by clicking on its image in the cooking display (i.e., only one of six options were displayed on the screen at a time - table setting or one of the five food items). Hence, more switching and much increased attention management were required. Will these changes affect the distinction between measures and the influence of practice on the differences between performers? These are the main questions of the second experiment.

2.2.2 Participants: Thirty-four participants were recruited. Their recruitment process and screening were identical to those described in Experiment 1. All participants were students of the Technion, with native Hebrew language level. They aged between 21 and 33 (mean age: 26.0); 42% of them were females. Two participants were excluded from the analysis, due to not completing all the training program in the required duration and schedule. Thus, 32 participants

were included in the analysis. After finishing the task, participants received 250 NIS (about 70 USD). Three participants, who received the highest combined standardized scores in a session, received 50 NIS bonuses for the first five sessions.

2.3 A BATTERY OF FIVE COGNITIVE TASKS

2.3.1 Tasks and design: This experiment was designed to investigate the effect of practice on individual differences in cognitive tests using the administration format of cognitive test batteries. A battery was constructed with five commonly used cognitive tasks. It was administered and replicated in six sessions, representing practice.

The selected tasks are generally less complex than the Breakfast Task, but the format and structure of each combines several elements. How would performer differences in tasks be affected across the six replications? In addition, to examine the relationship between the cognitive tasks and Breakfast Task performance, two sessions of the six-screen version of the task were administrated one before and one after the six replication sessions of the cognitive task battery.

The first selected task for this experiment was the ANT developed by Posner (Posner and Rothbart, 2007)[37]. Although it is not a common member of cognitive test batteries, it is important because despite being a simple and discrete task, it was specifically constructed to separate the effects of exogenous and endogenous attention systems on single reaction times. This separation of effects is like the observed distinction in the performance of the Breakfast Task between the local monitoring food accuracy factor (exogenous attention) and the executive control factor of table setting and synchronous serving (endogenous attention). Will the different ANT measures preserve their distinctiveness following practice?

Two additional tasks are working memory tasks, the **N-back** representing verbal working memory (Kane, Conway Miura, Colflesh, 2007[38]. Jaeggi, Buschkuehl, Perrig, Meier, 2010[39]) and the **Corsi Block Tapping** representing visual spatial working memory (Kessels Zandvoort Postma Kappelle Haan,2000[40], Arbuthnott & Frank, 2010[41]). Both are widely used in cognitive test batteries and experimental research. In neuropsychology, Corsi has been linked to the right hemisphere specialization and the N-back to the left hemisphere (Kessel et. al. 2000[40]). The two types were also separated in Baddely's (2012)[42] model of working memory (WM). Nonetheless, there has been accumulated research that shows that not only working memory load, but additional variables influence performance in the two tasks. For N-back, those variables include familiarity with the erroneous stimulus in the sequence (Kane et. al. 2007[38]) and the included processes of decision, selection, inhibition, and interference resolution (Jaeggi et. al. 2010[39]). Corsi does not only distinguished by being a measure of visuospatial rather than verbal memory but also relies heavily on memory for temporal information (Kessel et. all. 2000[40]). For both tasks, these variables reflect the features and format in which the tasks were structured and administered to evaluate the targeted WM ability.

Trails is the fourth selected task, including the A and B versions of the task. Trails A screen contains only single digits or letters presentations. Participants are requested to draw a line connecting the first to the last value stimulus, scattered on the screen. Trails B screens display both digits and letters and participants are requested to draw a connecting line, alternating between digits and letters in order. Both Trails versions were shown to equally reflect perceptual speed and fluid intelligence (Salthouse 2005[13]). Trails B performance was also shown to include the additional costs of executive control and task switching (Salthouse 2011[43]; Sanchez-Cubillo et. Al[44]. 2009; Arbuthnott & Frank, 2000[41]). Thus, while performance of both A and B task versions is associated with the same basic processing demands, new elements and format of the B version introduce new demands that influence performance.

The fifth task in the cognitive battery is the **Digit Symbol Substitution Task** (DSST). It has been adopted from the Wechsler intelligence battery. It is a paired association task, and performers are required to match digits with associated geometric symbols. The task was shown to be sensitive to impairments and improvement in processing speed (Hoyer et. al., 2004[45]). DSST performance was shown to correlate with real-world functional outcomes (e.g, the ability to accomplish everyday tasks) and recovery from functional disability. In addition, the DSST has been demonstrated to be sensitive to change in cognitive functioning in patients with MDD and may offer an effective means to detect clinically relevant treatment (Jaeger, 2018[46]). A detailed description of the 5 tasks is presented in Supplementary material A.

Supplementary material A: Cognitive Battery tasks
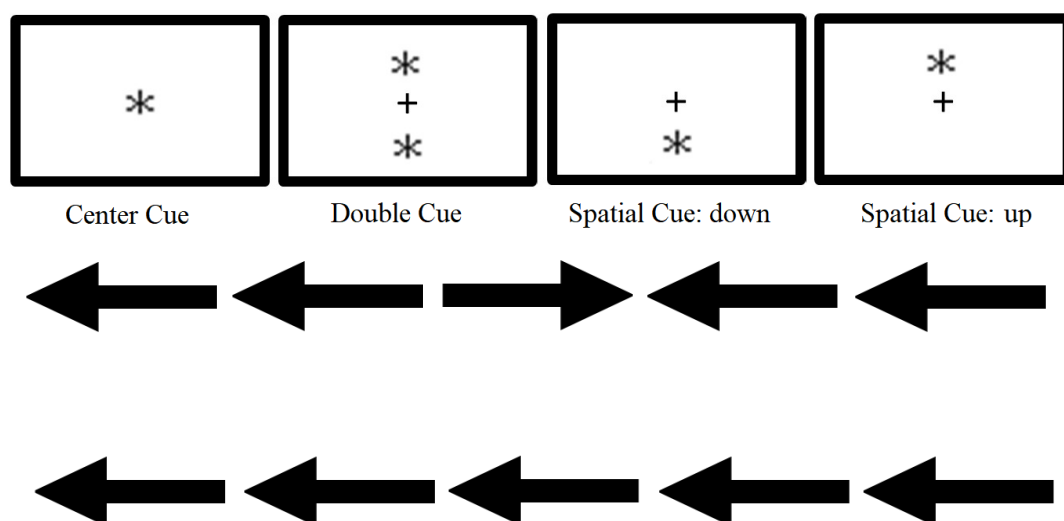The cognitive battery included five tasks replicating six times (sessions), in the same format and order. All five tasks were remotely accessed by the participants on their PCs at home, using a downloaded Millisecond Inquisit ® psychological tasks, on-line library (Inquisit 5, 2016)[52]. The five tasks were translated into Hebrew (stimuli and instructions) and shown in fixed order as in their short description below.

Attention Network Task (ANT): The ANT task was specifically designed to distinguish and study the cost and interrelations between exogenous (alert, orient) and endogenous (conflict resolution) attention systems (Posner & Rothbart, 2007[37]). It was selected because of its distinction between exogenous and endogenous categories of attention demands, which are closely associated with the local monitoring and executive control demand types of the Breakfast Task. It has been widely studied and modeled to support the distinction between exogenous and endogenous systems (Posner & Rothbart, 2007[37]; Peterson and Posner, 2012)[53] It is a computerized task. It displays a central arrow flanked with arrows indicating direction with congruent and incongruent arrows compared to the central arrow. The participant needs to indicate whether the central target arrow is pointing rightward or leftward (and typing "I" or "E", respectively). Participants were told that the task measured attention and were instructed to perform as fast and as accurately as they could. They were informed about the roles of cues and flankers and were encouraged to use them. The task aim was to capture three attention dimensions following Posner's attention model (MacLeod et al., 2010)[54]. The calculations of these three are based on differences of experimental conditions varying in the congruence of flanker arrows and spatial cues. The *alert effect* is calculated by mean response time of

no cue conditions minus the response time of double cue conditions; *Orienting effect* is calculated by mean response time (RT) of central cue trials minus mean RT of spatial (up and down) cue conditions; the *conflict effect* was calculated by mean RT of all incongruent flanker conditions minus all congruent flanker conditions. The Cues and flankers are shown in Figure 6. In addition to them, no cue condition and no flankers condition (only a single arrow) existed also (Fan et al., 2002)[55].

Although simpler and less demanding than the Breakfast Task, performance on all tasks of the cognitive battery was shown to be influenced by variables that reflect the format in which the tasks were structured and administered, to represent the targeted cognitive ability. Will such an influence also be observed in the effects of practice? The results obtained in the two experiments with the Breakfast Task demonstrated that the dominance and differential effect of practice depend on the nature and basic demand of the specific subcomponent of the task, over and beyond manipulations of global task demands. Using these findings as an anchor point, we hypothesize that specific components within tasks may vary in demand and consequently incur a differential effect of practice on task performance and influence the differences between performers.

Figure 3: *Cues and Flankers in the Attention Network Task*

Scatterplot of Intercept (x-axis) and Slope (y-axis) for RST, with Computed Regression with Negative Slope, Using Mixed Model Estimates for the 30 Subjects.

Each stimulus consisted of 2,000 Ms. inter stimulus interval (ISI) 100 Ms. cue, 400 MS interval between cue and target, and 1700 MS. target (or less if answered within time). Following a practice trial of 24 blocks with feedback, the task consisted of three blocks of 96 trials each with no feedback. Each trial was composed of two repetitions × four cue conditions × three flanker conditions × two target positions × two target directions. Within each 96 trials block, trials were distributed randomly (Fan et al., 2002)[55].

Trail Making task – The trail task has been employed as an indicator of processing speed and fluid intelligence (Salthouse, 2005[13], 2011[43]). Two versions of the task were given. The first (" *Trail A*"), included only numbers in ascending order, in which participants were required to draw a line connecting the first circle (numbered "1") to the last (numbered "25"). Circles were scattered around the screen. The second Trail version (" *Trail* B") included both numbers and letters in ascending order and alternately: participants were requested to draw a connecting line, alternating between numbers and letters in order. 1- a -2 - b - 3 – c, etc. Both Trail test versions included 25 circles (numbers or numbers and letters, with Trail B ending on "13"; Gaudino et al., 1995)[56]. Before each of the trails, there were short sample trials for training; participants were instructed to perform them as fast as they can. The program indicated when participants erred and coerced them to resume from the last correct step.

The number-letter version is considered to require more cognitive effort than the numbers only version requiring visual search, motor control as well as inhibition and executive control (Arbuthnott & Frank, 2010[41]). The trail task was computerized, and the paths were drawn by a mouse hovering movement through the circles and between them.

Corsi Block-tapping task is a computerized task assessing capacity of visuospatial working memory (visual memory span). The participants saw nine boxes in fixed locations on which a sequence is created by lighting different boxes one by one, in increasing sequence length (two different sequences for each span). The maximal sequence possible was of nine boxes; the task was adaptive, so after failing to recall a sequence twice it was terminated. Participants were asked to recall each sequence and repeat it immediately after it's done. The task is commonly used and is considered as the visual analogue for the digit span task (Kessels et al., 2000)[40].

N-back task is a common, demanding, phonological working memory task including aspects of executive control in which participants are required to decide for every letter presented in a sequence whether it is the same or different from the letter presented in the sequence N letters before. The value of N may refer to the previous letter (as in N = 1), 2 (as in N = 2) or even 3 letters before the letter (as in N = 3). For instance, in a N = 3 sequence condition: D, N, Y, D, K, Y, …, one should answer "yes" for the second D and the last Y (both in **bold**) because three items before them an identical letter was presented (both with underline). Zero value of N refers to the very first letter in the sequence (so it is fixed for each N = 0 sequence, thus any time the letter was presented along the sequence the participants should respond positively). There were three blocks for each level of N (0, 1, 2, or 3) with a 20 letters sequence per block. (Kane et al., 2007)[38]. Each letter was presented for 500 Ms., and the ISI was 2,500 Ms. Participants were asked to click "A" when the target was shown. The task yielded scores for number of hits, number of false alarms and discrimination value (hits minus false alarms, divided by number of blocks).

Digit Symbol substitution task is a part of the Wechsler intelligence battery (WAIS). Nine digits (1 – 9) are substituted by respective nine fixed meaningless symbols (using an injective function). Using a conversion key visible along the task, participants were required to replace, as fast and as accurately as they could, a sequence of digits, by the nine corresponding symbols. The conversion key (a table with the nine digits and the nine symbols underneath them) was presented in a designated box along the task (Wechsler, 1997)[57]. In the computerized task, the symbols switched roles with the digits, so

digits were to be typed on keyboard under their corresponding symbols, as accurately and as fast as possible within a limit of 120 seconds (same time limit as in the WAIS). The actual task was preceded by a short training on the conversion key.

2.3.2 Participants: Thirty-four participants were assigned to this experiment. All recruitment and screening processes were as reported in Experiment 2, the mean age was 26.0, and 42% of them were females. From among the 34-participants two were not included in the cognitive battery analysis as fast (thus: n = 32 for this part, and 31 participants for the N-back task – see below). For Breakfast sessions analysis: 32 participants were included. Two subjects were removed due to very poor performance, or not complying with the frequency demands of the practice program. There were thus 30 subjects who participated in both parts – cognitive sessions and Breakfast sessions. Participants received 200 NIS for their participation, with a possible 50 NIS bonus for the best 10% performers.

2.4 STATISTICAL ANALYSIS

Examination and statistical analysis of the performance results obtained, focused on three main questions:

1. Were there significant individual differences in the three performance measures of the Breakfast Task?

2. Were there significant effects of the practice sessions?

3. Was there a consistent relationship between individual differences and practice?

We used the R statistical package, and particularly the *lme4* package, to analyze the data and answer the three questions. The analysis uses all data points obtained by each study subject, so as to enable discerning whether individual intercepts and slopes differ significantly (mixed model) or whether they are essentially the same for all individuals (fixed model). In the present case the data consisted of three performance scores obtained for each of the eight rounds (trials) of the Breakfast Task over the five practice sessions.

The analyses used to answer the three questions were based on considering two linear regression models: one consisting of only *fixed effects* (**fixed model**) and one with both fixed and random effects (**mixed model**). *Individual differences* are considered to be part of the noise or error in the fixed model, whereas they are treated as *random effect*s in the mixed model. As we are considering linear regression models, the performance y_{ijk} of subject , on $trial\ k$ of session j can be expressed as a linear function plus error: $y_{ijk} = \alpha_i + \beta_i j + \epsilon_{ijk}$. For the fixed model, when no underlying individual differences are considered, $\alpha_i = \alpha_0$ for all i and $\beta_i = \beta_0$ for all i – meaning each subject has the same underlying intercept α_0 and slope β_0. Note that the slope measures the effect of practice as the subject proceeds from session to session (j runs from 1 through 5).

In the mixed model analysis, each subject has their own slope and intercept which are considered to be drawn from a bivariate normal distribution with mean $(\alpha_0, \beta_0)^T$ and variance-covariance matrix given by variances σ_α^2 and σ_β^2 on the diagonal and covariance $\rho\sigma_\alpha\sigma_\beta$. Here ρ represents the correlation coefficient between an individual's intercept and practice slope. If the variances are zero (0), then the random effects are constant, and the model reduces to the fixed model – which corresponds to the assertion that there are no individual differences between subjects: they all share the same underlying intercept and slope. Turning now to the three specific questions we proceed as follows.

1. **Individual differences**: For testing whether individual differences are important we compare the *fixed model* (which has $\sigma_\alpha^2 = 0 = \sigma_\beta^2 = \rho$) to the *mixed model* using a χ^2 likelihood ratio test based on 3 degrees of freedom. If the statistics are significant, it implies that individual differences should be accounted for in the analyses, and therefore further analyses should be based on the mixed model.

2. **Practice:** The slope β_0 expresses the average effect of practice per session. We expect it to be positive for measures that increase with better

performance (such as the number of set tables) and negative for measures that decrease with better performance (such as time to complete). *The estimate of the average slope is a fixed effect of the mixed model,* and is tested using a t-test, where the null hypothesis is H0: $\beta_0 = 0$ –that is, that there is no average effect of practice (or no learning). If we reject H0, we can claim that there is, on *average* over subjects, an effect of practice. The degrees of freedom for the t-test use the Satterthwaite method.

3. **Relationship between individual differences and practice**: The correlation coefficient ρ indicates the strength and direction of the relationship between an *individual's intercept and their slope*. For negative values of ρ, the relationship is such that subjects with larger intercepts tend to have smaller slopes, and vice versa. It is even possible that some subjects with large intercepts have negative slopes while others with small intercepts have positive slopes. In such circumstances, the *average* slope (β_0) may not be significantly different from zero. To test the significance of the correlation versus H0: $\rho = 0$, we fit a mixed model assuming independent slopes and intercepts and compare it to the full mixed model using the likelihood ratio test and a χ^2 statistic on one degree of freedom.

3. Results

BREAKFAST TASK (ALTERNATING SCREENS)
The results are presented considering the three questions. Note that observations on an individual trial that had values 2.5 standard deviations worse than the average for a given session were removed as outliers. Less than 2.5% of data were considered outliers.

Table 1: Averages and Standard Deviations of Breakfast Task Performance in the 5 Sessions (N = 30).

	Session 1	Session 2	Session 3	Session 4	Session 5
RST					
	13.31	6.57	6.71	5.12	6.35
	(10.37)	(5.91)	(6.38)	(4.97)	(6.38)
DIS					
	1.90	1.93	2.74	2.67	2.05
	(1.24)	(1.24)	(3.12)	(4.00)	(1.87
TS					
	7.34	8.05	8.32	9.17	9.32
	(1.09)	(1.57)	(1.99)	(1.51)	(1.70)

Note. Averages and SDs of Cooking Discrepancy (DIS), range of stop times, (RST = inaccuracy of synchronous serving time), Table setting (TS=number of correct tables set). SDs calculated from subject averages (over trials) are shown in parentheses

INDIVIDUAL DIFFERENCES
The comparative fixed and mixed analysis of results showed significant individual differences in all three performances measures: Discrepancy ($\chi_{(3)}^2 = 104.45$, p < .001), RST ($\chi_{(3)}^2 = 283.77$, p < .001) Table Setting ($\chi_{(3)}^2 = 651.06$, p < .001).

PRACTICE EFFECTS
Discrepancy: no significant average practice effects (t (41.4) = 0.850, p = .400). RST: significant practice (session) effects (t(29.8) = -4.56, p < .001). Table setting: a very significant practice effects (t)29.8) = 10.68, p < .001).

CORRELATION BETWEEN INITIAL ABILITY AND PRACTICE

The correlation between subjects' slope of practice and intercepts were: Discrepancy: not significant ($\chi_{(1)}^2 = 1.58, p = 0.21$); RST: (r=-.87), highly significant negative ($\chi_{(1)}^2 = 25.3, p < .001$). Table Setting: not significant ($\chi_{(1)}^2 = 1.69, p = 0.194$).

Figure 2 depicts the RST negative slope values of the 30 subjects. Initial lower score subjects gained more from practice than the initial higher scores.

3.2 BREAKFAST TASK (SIX SCREENS)

Table 2 presents the averages and standard deviation in the six sessions of the Breakfast Task performance.

Figure 2: *Scatterplot of Intercept (x-axis) and Slope (y-axis) for RST, with Computed Regression with Negative Slope, Using Mixed Model Estimates for the 30 Subjects.*
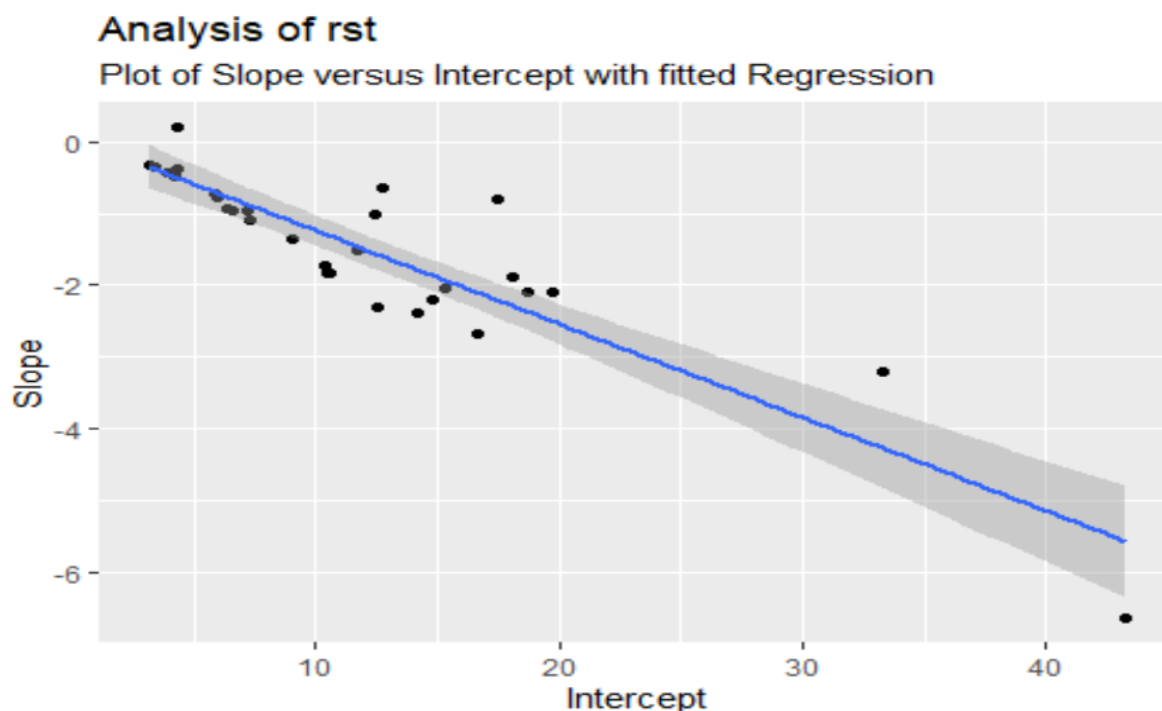

Table 2: Averages and Standard Deviations of Breakfast Task Performance (N=32).

	Session 1	Session 2	Session 3	Session 4	Session 5
RST					
	24.32	15.21	15.61	14.44	11.60
	(19.20)	(10.52)	(11.54)	(10.10)	(6.86)
DIS					
	3.83	4.1	3.72	2.77	4.28
	(2.85)	(3.11)	(3.06)	(2.13)	(3.49)
TS					
	7.46	8.38	9.07	9.55	9.67
	(1.74)	(2.02)	(2.06)	(2.52)	(2.54)

Note. Averages and SDs of Cooking Discrepancy (DIS), range of stop times, (RST = inaccuracy of synchronous serving time), Table setting (TS=number of correct tables set). SDs calculated from subject averages (over trials) are shown in parentheses.

Compared to performance averages in Experiment 1(Table 1), it can be seen that the two cooking error performance measures, averaged over the five sessions, were considerably higher and significantly worse, RST ($F(1,59) = 15.09$, $p<.001$, DIS ($F(1,59) = 5.10$, $p=.028$, $\eta p2 = .080$). On table setting, performance levels did not differ between the two experiments ($F (1,59) = 0.98$, $p=.327$).

INDIVIDUAL DIFFERENCES:
The comparative fixed and mixed analysis of results showed significant individual differences in all three performance measures:

DIS ($\chi_{((3))}^2 = 215.05$, $p < .001$), RST($\chi2(3) = 542.94$, $p < .001$), TS ($\chi2(3) = 1142.6$ $p < .001$).

Practice Effects

DIS, no significant practice effect was found ($t(32.0) = -0.305$, $p=.763$,).

RST, practice effects were significant ($t (32.1) = -4.385$, $p<.001$).

TS, practice effects were significant ($t (32.0) = 6.542$, $p<.001$).

Correlation between initial ability and practice

DIS, marginally significant negative correlation ($r = -0.50$) between subjects' intercept levels and practice slope, ($\chi2(1) = 3.095$, $p = 0.079$). RST, highly negative correlation ($r= -0.95$) between subjects' intercept and practice slope ($\chi2(1) = 57.38$, $p < .001$). Initial low performers gained more from practice than higher performers.

TS, no significant correlation between subjects' intercept levels and practice slope ($\chi2(1) = 1.158$, $p = .0.282$).

3.3 COGNITIVE TASKS BATTERY
We first present the results and analyses conducted on the cognitive tests battery. We then examine its relationship to breakfast task performance.

Table 5 presents the inter correlations between the tasks in the first and sixth replications, as well as between the performance of the same task in the first and last replication.

It can be seen, that although all correlations but one, between the first and the sixth replication of the same task are significant, two are significant only at the .05 level and five are significant at .01 level, with values ranging between .51 and .78, or between 26 – 61 percent of accounted variance (R2). With the six replications there was a considerable difference between the first and last session's task performance scores as indicators of the targeted cognitive function. The respective changes may reflect the differential influence of practice, reliability, fatigue, or a combination of these effects. The lowest and the only non-significant correlation was for the ANT conflict measure, which within the ANT task represents the executive control attention system. Taken together, the results indicate that the five tasks are relatively different from each other and show the general influence of replications. They justify an examination of the effect of practice on each test by itself.

As in the previous two experiments, a mixed linear model analysis was conducted for all tasks using the R package.

ANT -Alert: Very significant individual differences were found ($\chi2(3) = 126.97$, $p < .001$). Session effect was not significant ($t (32) = 0.62$, $p = .542$). However, correlation between intercept and slope was significant ($r = -.948$, $\chi2(1) = 5.98$, $p = 0.014$). In the Alert reaction time measure, better performance implies lower RT score. Hence, the high negative correlation between intercept and practice sessions indicates that performers with lower reaction time in the first session had a slower reaction time (less negative slope) to alerts along the six replication sessions while those who started with initial high reaction times improved more during the following six replications. Note that the negative correlation was also accompanied by a nonsignificant sessions effect.

Table 3: Averages and Standard Deviations (in parentheses) of each of the Cognitive Tasks in the Six Replications (N = 32).

	Session 2	Session 3	Session 4	Session 5	Session 6	Session 7
Alerting effect	48.88 (18.82)	54.88 (25.90)	49.87 (23.70)	47.44 (32.30)	52.52 (37.26)	57.76 (56.34)
Orienting effect	24.41 (14.43)	21.48 (21.63)	14.92 (18.72)	18.18 (16.59)	18.14 (29.27)	23.11 (26.71)
Conflict effect	73.23 (18.98)	62.28 (16.09)	52.84 (20.33)	53.91 (18.69)	52.36 (17.75)	45.53 (18.59)
Trail A	33.20 (7.89)	29.56 (8.43)	28.30 (9.07)	25.73 (7.53)	25.65 (7.23)	26.04 (8.62)
Trail B	40.99 (10.84)	35.35 (11.56)	34.98 (10.79)	29.97 (8.90)	29.55 (9.32)	28.21 (9.83)
Corsi Block	6.78 (1.13)	6.56 (1.11)	6.81 (1.20)	7.00 (1.32)	7.09 (1.00)	7.34 (1.04)
N back (n = 31)	53.77 (5.75)	55.39 (4.54)	54.84 (5.22)	55.03 (7.31)	54.29 (8.06)	53.61 (9.05)
DSS	76.47 (21.98)	95.53 (32.27)	114.00 (41.16)	125.97 (41.77)	129.78 (42.67)	139.88 (48.66)

Note. For ANT and Trail, lower time means better performance. For Block span, N-back, and Digit Symbol Substitution, higher counts better performance.

Table 4: Pearson Correlations Between Cognitive Tasks in First and Last sessions (S2 and S7) Separately.

S2 \ S7	Ale.	Ori.	Con.	T1	T2	CB	NB	DSS
Alerting	.42*	.61**	.16	.12	.11	-.14	-.74**	-.16
Orienting	.29	.51**	-.09	.01	-.01	-.20	-.52**	-.31
Conflict	.17	.27	.11	.31	.20	.10	-.34	-.26
Trail A	-.14	.13	-.06	.61**	.90**	-.28	-.36*	-.45**
Trail B	-.18	.21	-.36*	.64**	.52**	-.25	-.29	-.42*
Corsi Block	.17	.00	-.15	-.34	-.35*	.65**	.20	.21
N Back	.02	-.07	-.35	-.12	.01	-.11	.45*	.25
DSS	-.21	-.17	.03	-.36*	-.38*	.01	.30	.78**

*$p<.05$, **$p<.01$

Note. Lower-left triangle – Session 2; Upper-right triangle – Session 7. On the gray-shaded diagonal the correlation for each task between first and last sessions. Abbreviations on the heading row correspond to test names in the left column.

It can be seen, that although all correlations but one, between the first and the sixth replication of the same task are significant, two are significant only at the .05 level and five are significant at .01 level, with values ranging between .51 and .78, or between 26 – 61 percent of accounted variance (R2). With the six replications there was a considerable difference between the first and last session's task performance scores as indicators of the targeted cognitive function. The respective changes may reflect the differential influence of practice, reliability, fatigue, or a combination of these effects. The lowest and the only non-significant correlation was for the ANT conflict measure, which within the ANT task represents the executive control attention system. Taken together, the results indicate that the five tasks are relatively different from each other and show the general influence of replications. They justify an examination of the effect of practice on each test by itself.

As in the previous two experiments, a mixed linear model analysis was conducted for all tasks using the R package.

ANT -Alert: Very significant individual differences were found ($\chi 2(3)$ = 126.97, p < .001). Session effect was not significant (t (32) = 0.62, p = .542). However, correlation between intercept and slope was significant (r = -.948, $\chi 2(1)$ = 5.98, p = 0.014). In the Alert reaction time measure, better performance implies lower RT score. Hence, the high negative correlation between intercept and practice sessions indicates that performers with lower reaction time in the first session had a slower reaction time (less negative slope) to alerts along the six replication sessions while those who started with initial high reaction times improved more during the following six replications. Note that the negative correlation was also accompanied by a nonsignificant sessions effect.

Positive slope values imply that subjects with low initial response time (intercept), in the first session, increased their response time in the following sessions. The two lowest first session intercepts (best subjects), had very large positive slopes, The highest intercept (worst subject), had a very large negative slope. All the other, twenty-nine subjects, were distributed around the zero with very small positive or negative slopes.

ANT-orient: Very significant individual differences were found ($\chi 2(3)$ = 13.78, p = .003); session was not found significant (t(179.1) = -0.42, p = .674). Because the random intercept variance was estimated at 0, there was no correlation to be found between individual intercepts and slopes.

ANT-conflict: Very significant individual differences were found ($\chi 2(3)$ = 46.91, p < .001) and for the session effect (t (32) = -7.03, p < .001). Correlation between intercept and slope (r = -.642), was not significant ($\chi 2(1)$ = 1.13, p = .288).

Corsi Block Tapping: Very significant individual differences were found ($\chi 2(3)$ = 62.79, p < .001), and also for the average session effect (t(160) = 3.79, p < .001); Correlation between intercept and slope was not significant ($\chi 2(1)$ = 0.014, p = .90). The random slope variance was estimated close to 0, so there was no correlation to be found between individual intercepts and slopes (p=0.904).

N-Back: Very significant individual differences were obtained ($\chi 2(3)$ = 175.59, p < .001), but the session effect was not significant (t(31) = -0.398, p = .693). The correlation between intercept and slope was negative (r = - .52) and significant ($\chi 2(1)$ = 5.23, p = .022).

Trail A: Very significant individual differences were found ($\chi 2(3)$ = 157.7, p < .001), and also session effect (t (32) = -6.44, p < .001). Practice effects were similar for first session low and high performers and the negative correlation (r = -.47) between intercept level (first session) and slope was not significant ($\chi 2(1)$ = 2.01. p = .156).

Trail B: Very significant individual differences were found ($\chi 2(3)$ = 148.8, p < .001), and for the session effect (t (31) = -7.58, p<.001); Unlike Trail 1, the negative correlation between intercept levels and slope was significant (r = -.72, $\chi 2(1)$ = 9.78, p = .002).

Digit Symbol Substitution: Very significant individual differences were found ($\chi2(3) = 292.08$, $p < .001$), and for session effect ($t (32) = 10.48$, $p < .001$). However, correlation between intercept and slope was not significant ($\chi2(1) = 0.14$, $p = .708$.

Breakfast Task performance: A single eight-trial session of the six screens version was performed before the cognitive task replications and once after them. There were high and significant correlations obtained between the two sessions for the three performance measures: DIS, RST and TS were all significant at $p<.01$ level.

When correlated with the battery of cognitive tasks, TS had the largest number of significant correlations with the tasks in both sessions, before and after the six replications (ANT-conflict, Trails A and B, N-back and Corsi Block Tapping). Cooking RST did not correlate with any cognitive task in the first or after six replications. DIS scores had only a few significant correlations in the first session with ANT- conflict, Trails A and B, but none following the six replications of the cognitive battery.

Discussion

The leading motivation for the present study has been the claim that in the common evaluation of individual differences in cognitive abilities is mostly based on single sessions and practice on tasks has been limited or not included in the current use of cognitive test batteries. We argue that in addition to representing its targeted cognitive ability, task format features and instructions include many elements which influence performance levels and may be irrelevant to the evaluation of targeted cognitive ability (e.g., working memory, processing speed, inhibition, task switching). Practice would serve as a common experience on a task to all its performers and may reveal the initial and later effects of such components. It may also reduce the effects of the first session on performance of specific elements arising from differences in past histories and experience.

Significant individual differences in the first session performance levels were obtained for all tasks and measures in the three experiments, supporting their use to differentiate between individuals in their embedded cognitive abilities. However, practice had differential effects on tasks and their performance. Three types of practice effects were observed: No significant effects on individual performance levels; Significant and equal gains for different first meeting performance levels; Interactive effects between practice gains and first session different performance levels. Individual differences and differential practice effects were revealed not only for global performance on a task but also for specific components within it. These effects are considered in reference to the format of a given task, its nature and attention demand.

The alternating screen Breakfast Task was presented as an integrated single mission of coping with multitasking demands, but the effects of practice on its three performance measures differed. On table setting voluntary investment measure, performance improved with equal gain for performers with different initial score levels. In cooking coping with imposed load, accuracy (DIS) performance did not improve with practice, whereas the executive control serving synchrony measure (RST) had a negative slope with practice and significant negative correlation between individual slopes and intercepts indicates that performers with poorer initial performance gained more from practice than higher initial performance subjects. These differences reflect also the distinction between the three measures proposed by Rose et. al. (2015)[14] and Gopher et. al. (2022)[47], showing that in a single complex task, individual differences in basic performance ablities followed the nature, structure and demand of specific task segments rather than those of the global task as an integrative entity. They reinstate the question of identifying in tasks the segments and the relevant component of the enclosed cognitive abilities.

Rose et al. (2015)[14] study which did not include practice distinguished between the local monitoring nature of the food accuracy measure and the two executive control performance measures of food serving synchrony (RST) and table setting (TS). Gopher

et. al. (2022)[47] further distinguished between executive control of voluntary investment and coping with imposed load. Five sessions of practice did not have a significant effect on cooking accuracy performance levels but were significant and differentially affected by the two types of executive control performance measures. When cooking and balancing accuracy and synchronized serving measures, performers with initial lower synchronization levels benefited more from practice, while similar difference was not obtained in table setting voluntary investment.

In the six-screen version of the Breakfast Task despite the global much higher load and elevated attention management requirements, the same distinction between the three measures of practice effects individual differences. The higher load and attention demand of the six screens affected cooking performance (Table 1and 2). Both cooking accuracy (DIS) and synchronous serving (RST) error averages were significantly increased, with the most pronounced being the impairment of RST performance. Table setting (TS) performance was not impaired in the six-screen condition, with no meaningful change from the alternating screens experiment. These findings are consistent with previous studies (Craik &Bialystok (2006)[36] (Gopher et. al. 2022)[47]. Table settings distinct voluntary and static nature enabled subjects to develop and apply proper strategies to cope with the increasing attention management and performance requirements.

The five tasks employed in cognitive battery differ in their targeted dimension of cognitive ability assessed. The effects of practice on the tasks were also not uniform. In all tasks individual differences, among initial performance levels (intercepts), practice effects (slopes) or both, were a highly relevant component of analysis. What is the information added by the differential effects of practice?

We first examine the differential practice effects on the performance of the ANT task which was theoretically constructed to separate and compare the behavioral operation of exogenous and endogenous attention systems (Posner and Rothbart, 2007). Our results show that they differ not only in reaction time costs during the first session but also in their practice effects (Table 4). In previous studies the exogenous attention system has been argued to capture attention and operate automatically. It has been shown to be fast but transient, whereas the operation of the endogenous attention system has been reported to be slower, more controlled and effortful (Bonder, & Gopher, 2019[48]; Carrasco 2011[49]; Kurtz et al., 2017[50]; Kahneman 2011[51]). Consistent with the distinction between the two attention systems, no practice effects on performance were obtained for the exogenous attention alert and orient cost measures, while a significant reduction of reaction time costs with practice were obtained for the endogenous attention conflict resolution measure, which is called upon to resolve the mismatch between the pointing direction of central and side presented arrows. The differential practice effects on the ANT task measures substantiate not only its proposed distinction between the operation modes of the two attention systems but also demonstrates that separate practice effects can be obtained for different components within a task.

The format and features of ANT task were intentionally developed to separate the cost and operation modes of attention systems but may also provide a conceptual framework for considering the differential practice effects observed in the performance of the Breakfast Task and the tasks included in the cognitive battery. In the Breakfast Task the cooking accuracy measure (DIS) is based on monitoring cooking time, in which performers monitor the time driven size changes in the five bar-graphs of food items (Figure 1). These changes are automatically captured by the exogenous attention system (see also Rose, 2015[14]). The synchronized food serving and table setting performance measures both require endogenous attention and executive control involvement. In the two Breakfast Task experiments no practice effects were obtained for the cooking accuracy performance measure (DIS), while significant practice effects were obtained for both the cooking serving synchrony (RST) and table setting (TS).

In both the ANT and the Breakfast tasks, practice effects were obtained for the within task components demanding endogenous attention operations and no effects for components automatically captured by exogenous attention. In comparison with ANT, the Breakfast Task was developed as a global indicator of attention management and multitask ability and is a much more complex, continuous and multi-element task. Nevertheless, in both tasks, practice effects depended on the nature and attention demands of the specific task components rather than those targeted by the global task constructed. This conclusion is also strengthened by the results of the second experiment with Breakfast Task, in which the same differential practice effects were maintained although the overall difficulty and attention management demands of the global task were much increased in the second experiment.

If practice effects on task performance underline the initial difficulties and differences between performers, it should be recognized that although significant individual differences were obtained in the first session performance for all three measures in each of the two tasks, practice effects were obtained only for those associated with endogenous attention, top-down operations. One important difference between the tasks that should be noted is that in the ANT task the costs of conflict resolution were significantly reduced with practice, but there was no interaction of practice with individuals first session performance levels. Similar results were obtained for table setting performance in the Breakfast Task. However, for the cooking synchronized food serving measure there was a significant negative interaction such that performers with lower performance levels in the first meeting gained more from practice than those with higher initial performance levels. Can the difference be associated with the structure and format of the task.

The ANT task depends on responses to discrete and short duration presentations and does not include a working memory component. The Breakfast Task is comprised of multiple segments and both table setting and cooking include working memory and

attention management demands. However, and like the ANT, table setting is under voluntary control static and discrete while cooking is a continuous and time dependent segment. While cooking, cooking duration of foods and their end points are both monitored by automatically following relative changes in the five food barographs.

In each cooking round of the Breakfast Task, the food serving synchrony measure in each round represents a subjective solution for a possible conflict between cooking accuracy and joint serving errors. The instructions emphasized their equal importance for scoring cooking performance. Once cooking has started, executive control directs a subjective decision for the relative weight of food accuracy and joint serving errors. This balancing and conflict resolution element in cooking is like arrows direction in the ANT task and does not exist in the table setting segment. The observed interaction may be an example for significant individual differences on multitask performance observed during the first meeting of cooking, but cross change with practice. Further research is needed to decide whether the observed difference reflects the targeted evaluation of multitask ability or is associated with other irrelevant factors.

The existence and impact of conflict resolution between two items captured by exogenous attention can be examined in two additional tasks which revealed significant interaction between practice and first meeting performance levels, the working memory N-back task and Trail-B, which called for ordered switching between digits and letters in creating trails between all items presented on the screen. In both tasks a significant interaction was found between the first session's performance levels and the practice slope. In both cases practice was more beneficial for lower than higher level performers.

In the N-back working memory task, a presented sequence of single digits or letters is attended by exogenous attention and executive control is called upon to compare the identity of an N-back item presently displayed. In the Trail B task version,

all relevant letters and digits are displayed on the screen and present in each step. Performers are asked for an ordered switch between digits and letters when they create a trail. In this task, attention control is called upon to choose the correct next step from the exogenous captured digits and letters. Thus, like the ANT cooking, top-down attention control operations are called upon to resolve conflicts created by automatically attended response alternatives.

Common to the three tasks is the call for endogenous executive control to resolve a conflict between two exogenous attention captured signals. The three tasks differ from the significant practice effects without interaction obtained in the performance of the ANT task. However, while the ANT task is discrete, based on response to single trials and have limited working memory demands, the Breakfast, N Back and Trails B tasks, each includes a heavy working memory component, continuous and time dependent or based on sequences on multiple presentations for each response. It is instructive that efficient employment of executive control in this later case create initial differences in performance levels which interact with practice.

The influence of the requirement to resolve exogenous attention created conflict can be further evaluated by the examination of practice effects on other tasks in the present study which include both working memory demands and multiple items presentation per response, but not exogenous attention conflict. Table setting in the Breakfast Task, Trails A task version, Corsi Block Tapping and the Digit Symbol Test, are all tasks which include a working memory and multiple items but not a created conflict. In all of them, individual differences and practice effects were significant but did not interact with first meeting performance levels. It appears that the need to resolve a conflict added an additional factor to the performed task, a requirement which increased its executive control difficulty, created differences in the first session performance levels and led to the cross interaction between the practice slope and first session intercept.

It is also interesting to examine the obtained practice effects in the context of the compensation and magnification training effect models that were described in the introduction. To recap, compensation models argued for larger training gains for initial lower performers (Karbach et.al, 2017[5]), while the magnification model reported higher comparative training gains in performers with initial high-performance levels (Foster et. al. 2017). In the context of the present study, we can examine the experimental tasks employed by the two models. In Karbach's "compensation" experiments, the focus was on training mixed multiple switching between two tasks to improve executive control. For example, a figure is presented in which one task was to decide whether it is fruit or vegetable and the second task to judge its size as large or small. In another experiment of displayed figures, subjects had to switch between judging planes/cars or single/pairs of objects presented. Subjects had to switch every two trials. The visual elements in each figure were automatically attended by exogenous attention and the conflict to resolve was which of the two categories was relevant. Thus, the tasks used by the proposed compensation model are very similar to the Trail B - digit and letter task, in which practice gains were larger for initial lower performers.

The magnification effects in the second model study reported the effects of training on working memory span, by teaching and practicing WM strategies such as mnemonics. A good example is the study by Foster et. al. (2017)[32]. In their running span WM training, subjects were instructed to remember the last X presented letters with the X number increasing in each level of difficulty. Single letters were then briefly flashed on the screen. The number of letters seen varied with each trial but was never fewer than the subjects were told to recall. It is an increasing difficulty WM task, but no conflict to resolve. In this condition, teaching a memory strategy benefits those with better initial WM performance. In its format, this task is like the Corsi Block Tapping task. In our experiments we did not observe a practice magnification effect, but we did not conduct specific

training to improve WM span. In all tasks in which working memory was called upon without a conflict to resolve, there were significant practice effects which did not distinguish between performers based on initial performance levels. Thus, in our experiments when practice effects are related to the nature of executive control required involvement and type of attention demands appear to be consistent with the results of the tasks used by the compensation and magnification training models.

Conclusions

The main findings and points of the three experiments can be summarized as follows Practice effects were not uniform and reflected the nature and processing demands of different tasks as well as components within task. These differences cannot be observed and evaluated in the results of first session's performance.

The differential effects on components within tasks imply that not a single but sperate scores should be given to individual performance levels in many cognitive tasks.

For the variety and nature of the cognitive tasks that were employed in the present study three major processing requirements were identified to characterize the nature and demand of task components: nature of attention demands, conflicting or switching between inputs and high working memory requirement.

Limits and future research

There are limits and questions to the results and theoretical perspective proposed by the present study to be followed in future research.

1. **Increased sample size**: Although the focus of this study was on practice effects and the results of three experiments were consistent and significant, larger sample sizes than the 30 participants in each experiment, is commonly employed when the focus is the measurement of individual differences. This limit calls for a replication of the study with a larger number of participants.

2. **Contribution of task components and format**: a major outcome to be supported in continuing research is that task format and its main component requirements should be separately evaluated when measuring individual differences in cognitive abilities. The present tasks and their format should be considered as one example. The nature and dimensions of possible contributing components and their effect should be further investigated and justified. This strong claim has major implications for research and application and needs to be tested and examined in future research.

3. **Linkage to existing models** What is the relationship and linkage between the approach and results of the present study to the existing factor analysis and latent structure task grouping approaches?

4. **Coping with imposed load versus voluntary investment efforts**: Gopher et. al (2022)[47], demonstrated the importance of this distinction when coping with load and multitask demands The Breakfast Task table setting was the only performance measure in which performers had the complete freedom to decide when and how long to invest in performing this task segment. We believe that when coping with load the degrees of freedom for performers in strategy development is highly important but a neglected cognitive dimension, in which individual differences should be studied.

References:

1. Oren, C., Kennet-Cohen, T., Turvall, E. & Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. Psicothema, 2014, 26, 117-126.

2. Kleper, D. & Seka, N. (2017). General Ability or Distinct Scholastic Aptitudes? A Multidimensional Validity Analysis of a Psychometric Higher-Education Entrance Test. Journal of Applied Measurement. http://jampress.org/abst.htm

3. Gopher, D. (1982). A selective attention test as a predictor of success in flight training, Human Factors 24, 173-183.

4. Charles, B. K., & Florah, O. M. (2021). A Critical Review of Literature on Employment Selection Tests. Journal of Human Resource and Sustainability Studies, 9, 451-469.
https://doi.org/10.4236/jhrss.2021.93029

5. Friedman, N. & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex 86, 186-204.

6. Karbach, J., Könen, T., & Spengler, M. (2017). Who benefits the most? Individual differences in the transfer of executive control training across the lifespan. Journal of Cognitive Enhancement, 1, 394–405. https://doi.org/10.1007/s41465-017-0054-z

7. Salthouse, T. A., & Ferrer-Caja, E. (2003). What needs to be explained to account for age-related effects on multiple cognitive variables? Psychology and Aging, 18, 91–110.

8. Eich S.T, MacKay-Brandt A, Stern Y, Gopher D (2016). Age-Based Differences in Task Switching Are Moderated by Executive Control Demands. J Gerontology B: Psychol Sci Soc Sci.

9. Bialystok, E., Craik, F. I., & Stefurak, T. (2008). Planning and task management in Parkinson's disease: differential emphasis in dual-task performance. Journal of the International Neuropsychological Society, 14(2), 257-265.

10. Tanguay, A. N., Davidson, P. S. R., Nuñez, K. V. G., & Ferland, M. B. (2014). Cooking breakfast after a brain injury. Frontiers in Behavioral Neuroscience, 8, 272.

11. Folstein MF, Folstein SE, McHugh PR. (1975). 'Mini-Mental Stat": a practical method for grading the cognitive state of patients for the clinician. J Psychiatry Res.; 12:189–198.

12. Lucas J, Ivnik R, Smith G, et al. 1998; Mayo's older Americans normative studies: category fluency norms. J Clin Exp Neuropsychol. 20:194–200. 20.

13. Salthouse. T. A. Relations Between Cognitive Abilities and Measures of Executive Functioning. (2005) Neuropsychology, 19, 4, 532–545.

14. Rose, N. S., Luo, L., Bialystok, E., Hering, A., Lau, K., & Craik, F. I. M. (2015). Cognitive processes in the Breakfast Task: Planning and monitoring. Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 69(3), 252–263.
https://doi.org/10.1037/cep0000054

16. Kabat, M. H., Kane, H.M., Jefferson, L.R., A. L. & DiPino, R. K (2001). Construct Validity of Selected Automated Neuropsychological Assessment Metrics (ANAM) Battery Measures. The Clinical Neuropsychologist Volume 15, 498-507.

17. Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology, 41, 49-100

18. Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences in skill acquisition. Journal of Experimental Psychology: Applied, 6, 259 –290.

19. Salthouse, T. A. (2004). Localizing age-related individual differences in a hierarchical structure. Intelligence, 32, 541–561.

20. Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. Current Directions in Psychological Science, 21, 8-14.

21. Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., et al. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: a longitudinal twin study. Developmental Psychology, 52, 326-340.

22. Frederiksen, J.R. & White, B.Y. (1989). "An Approach to Training Based Upon Principal-Task Decomposition", Acta Psychologica, 71, 89-146

23. Gopher, D., Weil, M., & Siegel, D. (1989). "Practice under changing priorities: An approach to training of complex skills". Acta Psychologica, 71, 147-179.

24. Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology, 58*, 79–87.

25. Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Developmental Psychology, 50*, 304–315.

26. Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*, 978–990.

27. Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. *Experimental Aging Research, 34*, 188–219.

28. Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental Psychology, 37*, 715–730.

29. Konen, T. & Karbach, J. Analyzing differences in intervention-related changes, (2021). *Advances Methodsand Practices in Psychological Science*, 4,1, 1-19.

30. Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: negative age differences in a mnemonic skill are robust. *Developmental Psychology, 28*, 121–125.

31. Brehmer, Y., Li, S. C., Mueller, V., von Oertzen, T. V., & Lindenberger, U. (2007). Memory plasticity across the life span: uncovering children's latent potential. *Developmental Psychology, 43*, 465–477.

32. Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1677–1689. https://doi.org/10.1037/xlm0000426

33. Lindenberger, U., Kliegl, R., & Baltes, P. B. (1992). Professional expertise does not eliminate age differences in imagery-based performance during adulthood. *Psychology & Aging, 7*, 585–593.

34. Lövdén, M., Brehmer, Y., Li, S. C., & Lindenberger, U. (2012). Training- induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience, 6*, 141.

35. Verhaeghen, P., & Marcoen, A. (1996). 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, 11* (1), 164–178.

36. Craik, F. I., & Bialystok, E. (2006). Planning and task management in older adults: cooking breakfast. *Memory & Cognition*, *34*(6), 1236-1249.

37. Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Anu. Rev. Psychol., 58*, 1-23.

38. Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. Journal of Experimental psychology: learning, memory, and cognition, 33(3), 615.

39 Jaeggi, S. M., Buschkuehl, M., Perrig, W.J. & Meier, B. (2010). The concurrent validity of the *N*-back task as a working memory measure. Journal of Experimental Psychology: Learning, Memory, and Cognition 2007, Vol. 33, No. 3, 615– 62.

40. Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. Applied neuropsychology, 7(4), 252-258.

41. Arbuthnot, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. Journal of clinical and experimental neuropsychology, 22 (4), 518-528.

42. Baddeley Alan Working Memory: Theories, Models, and Controversies Annual Review of Psychology (2012)63, pp. 1 – 29

43. Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? Intelligence, 39(4): 222–232.

44. SÁNCHEZ-CUBILLO, J.A. PERIÁÑEZ, D. ADROVER-ROIG, J.M. RODRÍGUEZ-SÁNCHEZ, M. RÍOS-LAGO, J. TIRAPU, BARCELÓ F. (2009): Task-switching, working memory, inhibition/interference control, and visuomotor abilities. Journal of the International Neuropsychological Society.15, 438–450. doi:10.1017/S1355617709090626

45. Hoyer, W. J. Stawski, R. S. Wasylyshyn C, Verhaeghen P. (2004): Adult Age and Digit Symbol Substitution Performance: A Meta-Analysis. Psychology and Aging Inc. 19(1), 211–214.

46. Jaeger j. (2018): Digit Symbol Substitution Test. J Clin Psychopharmacol 38: 513–519.

47. Gopher D, Ben-Eliezer D, Levine A (2022). Imposed load versus voluntary investment: Executive control and attention management in dual-task performance. Acta Psychologica, 227 103591. doi.org/10.1016/j.actpsy.2022.103591

48. Bonder, T., & Gopher, D. (2019). The effect of confidence rating on a primary visual task. Frontiers in Psychology, 10, 2674.

49. Carrasco, M. (2011). Visual attention: The past 25 years. Vision research, 51(13), 1484-1525.

50. Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of Endogenous and Exogenous Spatial Attention on Decision Confidence. Scientific Reports, 7(1). https://doi.org/10.1038/s41598-017-06715-

51. Kahneman Daniel: (2011), Thinking fast and Slow, Penguin Books.

52. Inquisitr 5 [Computer software]. (2016). Retrieved from https://www.millisecond.com.

53. Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. Annual review of neuroscience, 35, 73-89.

54. MacLeod, J. W., Lawrence, M. A., McConnell, M. M., Eskes, G. A., Klein, R. M., & Shore, D. I. (2010). Appraising the ANT: Psychometric and theoretical considerations of the Attention Network Test. Neuropsychology, 24(5), 637.

55. Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. Journal Cognitive Neuroscience, 14(3), 340-7.

56. Gaudino, E. A., Geisler, M. W., & Squires, N. K. (1995). Construct validity in the Trail Making Test: what makes Part B harder? Journal of clinical and experimental neuropsychology, 17(4), 529-535.

57. Wechsler, D. (1997). Wechsler Adult Intelligence Scale--Third Edition (WAIS-III).