

Microbiota shaping: the effects of probiotics, prebiotics and fecal microbiota transplants on cognitive functions. A systematic review.

Simone Baldi, Tiziana Mundula, Giulia Nannini, Amedeo Amedei

Simone Baldi, Tiziana Mundula, Giulia Nannini, Amedeo Amedei, Department of Clinical and Experimental Medicine, University of Florence, Florence 50134, Italy

Amedeo Amedei, SOD of Interdisciplinary Internal Medicine, Azienda Ospedaliera Universitaria Careggi (AOUC), Florence 50134, Italy

Supplementary material

Summary of all cognitive tests performed in the selected papers

- *Aachener Aphasia Test (AAT)*

Is a method for diagnosing aphasia and it consists of spontaneous speech analysed for 6 parameters along a scale from 0 to 5. In particular, it evaluates communicative behavior, articulation and prosody, automatic language, semantic-lexical structure, phonemic structure and syntactic structure.

- *Allgemeine Depressionsskala (ADS)*

It is a 20 question standardized test that assesses the emotional, somatic and cognitive symptoms of depressive mood in the last week.

- *Attention and Perceptivity Test (APT)*

APT assesses selective attention, sustained attention, vigilance, visual perception and visual scanning. Test score is based on variables, such as work speed, number of errors, and number of omissions.

- *Attention Switching Task (AST)*

AST evaluates the participants' ability to switch attention between the direction or location of an arrow on screen. The test begins with an arrow in the centre of the screen, which points either to the left or to the right. The participant is introduced to two buttons, one on the left and one on the right, and they must either press the left or right button to indicate on which side of the screen the arrow is displayed, or else press the left or right button that correspond with the direction in which the arrow is pointing.

- *Auditory Continuous Performance Test (ACPT)*

The ACPT measures selective and sustained attention. Participants have to respond by pressing a button as soon as they hear the target stimulus.

Correct and incorrect detection, omission errors, and reaction time are measured while participants listen to random numbers in a time range and respond to the target stimulus.

- *Beck Depression Inventory (BDI)*

BDI is a self-reported questionnaire that consists of 21 items that assess the intensity of depressive symptoms using a Likert-like scale from 0 (symptom not present) to 3 (symptom very intensive). The total score ranges from 0 to 63, with higher scores indicating higher levels of depression.

- *Block Design Test (BDT)*

The subjects have to assemble blocks according to pre-specified patterns of increasing difficulty with time.

- *British Ability Scales (BAS-III)*

It is a standardised battery assessing a child's cognitive ability and educational achievement. It comprises 20 short tests, each measuring particular types of knowledge, thinking and/or skills.

- *California Verbal Learning Test (CVLT)*

CVLT measures episodic verbal learning and memory.

The experimenter reads a list of 16 nouns aloud, at one-second intervals, in fixed order, over five learning trials (list A). After each trial, the subject is asked to recall as many words as they can in any order (i.e., free recall). The words are drawn from four semantic categories (tools, fruits, clothing, spices, and herbs). An interference list (list B) is presented that shares two categories with List A (e.g., fruit and tools) and has two unshared categories (e.g., fish and kitchen utensils). Free and cued recalls of the list A are tested immediately (short-delay), and again after 20 minutes (long-delay). In cued recall, the experimenter prompts the subjects with the word category.

The CVLT ends with a recognition task, where the experimenter presents the subject with a 44-word list, and the subject must indicate whether it is a target word or a distractor. Some distractors share semantic categories with the target words while others sound alike.

- *Children Mood and Feelings Questionnaire (MFQ)*

Is a screening tool for screening of depression in children and it consists of a series of 33 descriptive phrases regarding how the subject has been feeling or acting recently.

- *CogState Brief Battery*

This battery of computer-based tests provides a measure of four core cognitive domains: processing speed, attention, visual learning and working memory by carrying out card tasks and maze tasks.

- *CogTrack™ System*

Is an online set of automated cognitive tests, which provide composite measures of attentional intensity, sustained attention, attentional fluctuation, cognitive reaction time, working memory capacity, memory retrieval speed and response variability.

- *Corsi Block Tapping Test (CBTT)*

It requires the subject to observe the sequence of blocks "tapped" (or lit up in the computer version), and then repeat the sequence back in order. The task starts with a small number of blocks and gradually increases in length up to nine blocks. The test measures both the number of correct sequences and the longest sequence remembered.

- *Depression, Anxiety and Stress Scale (DASS-42) questionnaire*

DASS-42 is a 42 item self-report validated inventory comprising of three scales design to measure the negative emotional states of depression, anxiety and stress, where each of the three scales contained 14 items. The depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia, and inertia. The anxiety scale evaluates autonomic arousal, skeletal muscle effects, situational anxiety, and subjective experience of anxious affect, while the stress scale assesses difficulty in relaxing, nervous arousal, and being easily upset/agitated, irritable/over-reactive and impatience. Subjects are assessed on a 4- point Likert scale (0 ¼ did not apply to me at all, 1 ¼ applied to me to some degree or some of the time, 2 ¼ applied to me to a considerable degree or a good part of the time, 3 ¼ applied to me very much or most of the time). Scores for each subscale are categorized into five severity ranges, namely normal, mild, moderate, severe and extremely severe

- *Digit Span Test (DS)*

It is used to evaluate working memory by asking participants to immediately recall a sequence of numbers. DS test is made up of two different tests:

1. Digits Forward (repetition of digits forward).
2. Digits Backward (repetition of digits backward).

The test consists of pairs of sequences of numbers; the examiner reads the numerical sequence (one number per second); when the sequence is correctly repeated by the subject, the examiner reads the next sequence, which is longer by a number than the previous one, and continues so until the subject fails a pair of sequences or correctly repeats the last sequence consisting of nine numbers.

- *Digit Symbol Test (DST)*

The subjects have to transcribe nonsense symbols into corresponding spaces over two minutes. This test evaluates psychomotor speed and working memory and the number of successfully transcribed is the score.

- *Emotional Decision Making (ED) and Emotional Recognition Memory (RM) tasks*

The ED tasks employed a block paradigm design with eight alternating activation (A) and baseline (B) blocks (BABABABA...). Five stimuli (7s each) are presented in each activation block preceded by a cue word (1s; N for neutral block, U for unpleasant block).

Stimuli are triplets of neutral or unpleasant pictures (either N or U) and the participants are instructed to select the most neutral or most unpleasant picture in the triplet, basing their decision on their personal experiences.

60 neutral and 60 unpleasant pictures are presented in the activation block. The baseline block is a matching task in which the stimuli (7s each) are triplets of geometric pictures and participants were instructed to select a picture with a fixation cross on it.

ER task was performed to assess the recognition memory for the emotional stimuli presented in the emotional decision-making task. 120 additional pictures (60 neutral and 60 unpleasant) were selected to serve as foils in the memory task. Subjects were instructed to press 'yes' or 'no' according to whether or not they recalled seeing the picture during the ED task.

One hour after the ED task, an ER task is performed to assess the recognition memory. 120 additional pictures (60 neutral and 60 unpleasant) are selected to serve as foils in the memory task. This task also employs a block design with 12 alternating active and baseline blocks. Subjects are instructed to press 'yes' or 'no' according to whether or not they recalled seeing the picture during the ED task. 10 stimuli (3s each) are presented for each activation block preceded by a cue word (3s; N for neutral block, U for unpleasant block).

During the baseline block, participants are presented with geometric pictures with or without fixation crosses and are required to press 'yes' if they saw a fixation cross and vice versa.

- *Emotional face-matching paradigm*

Stimuli are presented in a block design, with a total of 18 blocks consisting of three stimuli each. The task includes a control and an emotion condition. In the control condition, participants have to match one of two geometric shapes presented at the bottom, to a target shape shown at the top of the screen. The experimental condition involves participants choosing one of two emotional (angry or fearful) faces shown at the bottom of the screen that best match the emotional expression of a face seen at the top of the screen.

- *Emotional face-word stroop paradigm*

During this task, participants are presented with pictures of male faces expressing fear or happiness. On top of the faces, the words "happy" and "fearful" are presented in prominent red letters. The emotions described by the words are either congruent with the emotion of

the face or incongruent, and participants have to indicate the emotion of the face by ignoring the emotion word

- *Emotion Recognition Task (ERT)*

The Emotion Recognition Task is a computer-generated paradigm for measuring the recognition of six basic facial emotional expressions: anger, disgust, fear, happiness, sadness, and surprise.

Each face is displayed for 200ms and then immediately covered up to prevent residual processing of the image. The participant must select which emotion the face displayed from the six options.

- *Emotional Stroop task*

The emotional Stroop task is used as an information-processing approach to assessing emotions. Like the standard Stroop effect, the emotional Stroop test works by examining the response time of the participant to name the colors of the words presented to them. Unlike the traditional Stroop effect, the words presented either relate to specific emotional states or disorders, or they are neutral (e.g., "watch", "bottle", "sky").

- *Flanker task (modified)*

Five arrowheads are displayed per situation. The middle (target) arrowhead is surrounded by flanker arrowheads. In the congruent situation, all arrowheads point in the same direction as the middle arrowhead (<<<<< or >>>>>). In the incongruent condition, the middle arrowhead points in the opposite direction to that of the surrounding arrowheads (<<><< or >><>>). Using their index finger, participants are instructed to press an image on the screen as quickly as possible in response to the direction of the middle (target) arrowhead (left [>] or right [<]) and perform as many tasks as possible within 1 min.

- *Generalised Anxiety Disorder Questionnaire-7 (GAD-7)*

The GAD-7 includes seven items with four possible ranked responses each (not at all [score, 0]; several days [score, 1]; more than half the days [score, 2]; nearly every day [score, 3]). The total GAD-7 score ranges from 7 to 21 and is used as an anxiety index.

- *Hamilton Depression Rating (HAM-D 17)*

HAM-D 17 is a 17 item questionnaire used to provide an indication of depression. The severity of depression was assessed by a 3-point or 5-point Likert-type scale evaluating mood, feelings of guilt, suicide ideation, insomnia, agitation or retardation, anxiety, weight loss, and somatic symptoms.

- *Iowa Gambling Task (IGT)*

IGT is a psychological task thought to simulate real-life decision making.

Participants are presented with four virtual decks of cards on a computer screen. They are told that each deck holds cards that will either reward or penalize them using game money. The goal of the game is to win as much money as possible. The decks differ from each other in the balance of reward versus penalty cards. Thus, some decks are "bad decks", and other decks are "good decks", because some decks will tend to reward the player more often than other decks.

- *Leiden Index of Depression Severity (LEIDS)*

Is a self-report questionnaire assessing cognitive reactivity to sad mood or vulnerability to depression using different subscales: aggression, hopelessness, acceptance, control, risk aversion and ruminance.

- *Mini Mental State Evaluation (MMSE)*

The MMSE test is a test used to detect cognitive impairment. It evaluates different items grouped into five sections: orientation, immediate memory, attention and calculation, delayed recall, and language and construction. Each section is evaluated with a score range of 0–30 and the highest scores indicating better performance.

- *Montreal Cognitive Assessment instrument (MoCA)*

It represents a measure of general cognitive function and had a score range from 0 to 30, with higher scores indicating more optimal cognitive performance.

It assesses different cognitive domains such as attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation.

- *Motor Screening Test (MOT)*

This test evaluates sensorimotor skills. Coloured crosses are shown in different locations on the screen for 2 minutes, one at a time. The participant must select the cross on the screen as quickly and accurately as possible.

- *Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS)*

It is the most used rating scale assessing the prognosis of Parkinson's disease

The evaluation includes six parts: the clinical evaluation of mental status, behavior and mood, the ability to perform daily activities, clinical evaluation of motor skills, the clinical analysis of disease complications, the prognostic evaluation of the progress and severity of the disease according to the Hoehn and Yahr scale, the assessment of disability according to the Schwab and England autonomy scale

- *National Adult Reading Test (NART)*

A series of words, of decreasing frequency in English, are pronounced in a non-standard way, are presented in a written form. The number that can be correctly pronounced predicts measures of intelligence.

- *Number Connection Tests A (NCT-A) and B (NCT-B)*

NCT-A evaluates psychomotor speed while connecting dots from 1 through 25 while during NCT-B, subjects need to connect alternating numbers and letters. Time required to complete these correctly is the outcome for both NCT-A and NCT-B.

- *Paired Associates Learning (PAL)*

PAL assesses visual memory and new learning.

Boxes are displayed on the screen and are “opened” in a randomised order. One or more of them will contain a pattern. The patterns are then displayed in the middle of the screen, one at a time and the participant must select the box in which the pattern was originally located. If the participant makes an error, the boxes are opened in sequence again to remind the participant of the location of the patterns.

- *Patient Health Questionnaire-9 (PHQ-9)*

The PHQ-9 comprises nine items with four possible ranked responses each (e.g. question: over the past 2 weeks, how often have you been bothered by the following problems? Feeling nervous, etc.; possible responses: not at all [score, 0]; several days [score, 1]; more than half the days [score, 2]; nearly every day [score, 3]).

The total PHQ-9 score range from 9 to 27 and is used as a depression index.

- *Perceived Stress Scale-10 (PSS-10)*

PSS-10 is a ten-item self-reported questionnaire assessing the degree in which recent life situations are appraised as stressful. People have to indicate on a five-point scale ranging from 0 (never) to 4 (very often) how often they have felt or thought in a certain way during the period of the past month. All responses are then summed to measure the level of perceived stress.

- *Phonological Verbal Fluency (PVF) test*

The examinees are asked to orally produce words beginning with a given letter. F, A and S are the most commonly described in literature and the most commonly used in clinical practice for the rehabilitation of patients suffering from acquired neurological disorders

- *Positive and Negative Affect Schedule (PANAS)*

PANAS is a self-report questionnaire that consists of two 10-item scales to measure both positive and negative affect. Each item is rated on a 5-point scale of 1 (not at all) to 5 (very much).

- *Profile of Mood States (POMS)*

It is a psychological rating scale used to assess transient, distinct mood states.

The tests consisted of 72 adjectives to which the subjects responded using a Likert scale by indicating the extent to which each adjective had applied to them in the last week. The test measures six basic dimensions of mood; energetic/tired, clearheaded/muddled, composed/anxious, confident/unsure, elated/depressed and agreeable/angry. These basic mood dimensions were also rated daily; every evening subjects rated their mood throughout the day on six 10cm visual analogue scales with these pairs of adjectives at either end.

- *Psychometric Hepatic Encephalopathy Score (PHES)*

PHES represents the gold standard for the diagnosis of hepatic encephalopathy. It is composed of five tests, namely number connection, number connection test-B, serial dotting test SDT, line tracing test and digit symbol test and it can be used to assess motor speed, motor accuracy, concentration, attention, visual perception, visual-spatial orientation, visual construction and memory.

- *Rapid Visual Information-Processing (RVIP) task*

In the RVIP task, participants have to monitor a continuous series of digits for targets of three consecutive odd or three consecutive even digits. The digits are presented on a computer screen at a rate of 100 per minute in pseudorandom order, and participants respond to the detection of a target string by pressing the space bar quickly as possible. The task were continuous and lasted for 5 min, with eight correct target strings being presented during each minute.

The task is scored for the number of target strings correctly detected, average reaction time for correct detections, and number of incorrect responses. It takes approximately 1 h to complete the cognitive fatigue test.

- *Raven's Standard Progressive Matrices (SPM)*

It is a 60-item nonverbal group test measuring abstract reasoning and problem solving capabilities. In each test item, the subject is asked to identify the missing element that completes a pattern. Many patterns are presented in the form of a 6×6, 4×4, 3×3, or 2×2 matrix

- *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)*

This test includes 12 standard cognitive subtests that yield a total score and five index scores grouped into the following subscales: immediate memory (list learning and story memory), visuospatial/constructional (figure copy and line orientation), language (picture naming and semantic fluency), attention (digit span and digit symbol coding) and delayed memory (list recall, list recognition, story recall and figure recall). The raw scores of the 12 cognitive subtests were converted to age-based subset index scores for the five cognitive

subscales. The total score was calculated using the combination of the raw scores of the 12 subtests.

- *Retrieval from long-term memory*

The ability to recall the capital cities of a number of countries provided a measure of the ability to retrieve names from long-term memory. A list of 30 countries was given.

With cases where no answer was produced, subjects were asked to rate on a six-point scale whether they had the feeling of either knowing or not knowing the answer they could not recall

- *Rey Auditory Verbal Learning Test (RAVLT)*

It is a neuropsychological assessment designed to evaluate verbal memory and it can be used to evaluate the nature and severity of memory dysfunction and to track changes in memory function over time. The test is designed as a list-learning paradigm in which the patient hears a list of 15 nouns and is asked to recall as many words from the list as possible. After five repetitions of free recall, a second “interference” list (List B) is presented in the same manner, and the participant is asked to recall as many words from List B as possible. After the interference trial, the participant is immediately asked to recall the words from List A, which she or he heard five times previously. After a 20 min delay, the participant is asked to again recall the words from List A

- *Rey-Osterrieth Complex Figure (ROCF) test*

Is a neuropsychological assessment in which examinees are asked to reproduce a complicated line drawing, first by copying it freehand (recognition), and then drawing from memory (recall). Many different cognitive abilities are needed for a correct performance, and the test therefore permits the evaluation of visuospatial abilities, memory, attention, planning, working memory and executive functions.

Immediate recall: after a short delay, the examinee is asked to reproduce the figure from memory.

Delayed recall: after a longer delay (20–30 minutes), the examinee may again be asked to draw the figure from memory.

- *Ruff Figural Fluency Test (RFFT)*

RFFT is a non-verbal fluency test that measures executive functions, such as planning strategies, divergent thinking and the ability to shift between different cognitive tasks and executive ability to coordinate those processes.

The test consists of five 60-second parts, each with a different stimulus presentation.

The respondent draws as many unique designs as possible within 60 seconds by connecting the dots in different patterns. The total number of unique designs drawn constitutes the main score.

- *Semantic Verbal Fluency (SVF) test*

Participants have to produce as many words as possible from a semantic category in a given time (usually 60 seconds). This category can be semantic, including objects such as animals or fruits, or phonemic, including words beginning with a specified letter.

- *Socially Evaluated Cold Pressor Test (SECPT)*

During this test, physical and psychological stress was induced. Physical stress was induced by having participants immerse their hands into ice water (ranging between 0 and 3 °C) for “as long as possible, until the researcher indicates to pull the hand out of the water”. Unknown to the participants, the maximal duration was set at 3 min. The test was conducted by a researcher who was yet unknown to the participant, and who adopted neutral and socially distant behaviour to increase psychological stress. Further, psychological stress was induced by asking participants to look into a video camera during the cold-water test, with the aim to record their facial expressions.

- *State-Trait Anxiety Inventory for Children (STAIC)*

It consists of two twenty-item scales for measuring anxiety in children. It distinguishes between a general proneness to anxious behaviour rooted in the personality and anxiety as a fleeting emotional state.

- *Story Recall Test (SRT)*

It is a neuropsychological assessment for evaluating verbal memory function. It requires an individual to recall details of a story that is told or read to him or her. Five stories were found to be the minimum needed for reliability

- *Stroop Color and Word Test (SCWT)*

The Stroop effect permit to measure a person's selective attention capacity and skills, as well as their processing speed ability.

In the first trial, the written color name differs from the color ink it is printed in, and the participant must say the written word. In the second trial, the participant must name the ink color instead.

- *Symptom Checklist 90 (SCL-90)*

It is a self-report questionnaire which consists of 90 items, yielding nine scores along primary symptom dimensions and three scores among global distress indices. The primary symptom dimensions that are assessed are somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychoticism. The three indices are global wellness index, hardiness, and symptom free.

- *Test of Time and Weights Estimation (STEP)*

This test evaluates the abstraction skills and it requires that subjects estimate time intervals and object weights and it includes items for which individuals have either active or passive experience.

- *Test Your Memory (TYM)*

TYM is a cognitive test designed to detect Alzheimer disease.

It includes a series of 10 cognitive tasks, namely orientation, ability to copy a sentence, semantic knowledge, calculation, verbal fluency, similarities, naming, visuospatial abilities, ability to do the test and recall of a copied sentence giving a possible total of 50 scores.

- *Trail Making Test A (TMT-A) and B (TMT-B)*

Both parts of the Trail Making Test consist of 25 circles distributed over a sheet of paper. In Part A, the circles are numbered 1 - 25, and the patient should draw lines to connect the numbers in ascending order. In Part B, the circles include both numbers (1 - 13) and letters (A - L); as in Part A, the patient draws lines to connect the circles in an ascending pattern, but with the added task of alternating between the numbers and letters (i.e., 1-A-2-B-3-C, etc.). The patient should be instructed to connect the circles as quickly as possible, without lifting the pen or pencil from the paper

- *Two-choice task*

The two-choice task measures the tendency to choose a small reward over a larger delayed reward. For example, participants are presented with 40 trials in which they choose to wait 5 s for five points or 15 s for fifteen points. On each trial, a diamond and a star appear on the screen. Participants are required to select one of the two shapes. The length of the delay and the size of the reward associated with each shape remain constant across the testing session; for example, the star allow the participant to earn 5 points after waiting 5 s, while the diamond is followed by 15 points after waiting for 15 s. Impulsive choices are defined as the total number of short-delay reward responses.

- *Verbal Fluency Test (VFT)*

Tests of verbal fluency are used to measure the functioning of the frontal lobes of the brain. One of three letters will be given in turn, with the instruction to write down as many words as possible in 1 min beginning with that letter.

- *Visual Search Test (VST)*

Is a test in which people are asked to find a visual stimulus amongst other visual stimuli (distractors). The more stimuli there are, the more time it takes to find a target stimulus

- *Wechsler Memory Scale (WMS)*

Three stories was constructed of the same length with a similar number of items to be remembered. Each story was played on a tape recorder. Immediately after hearing the story, the subjects wrote down as many details as they could recall. At the end of the test

session, a measure of delayed recall was obtained by asking them a second time to write down the story.