4. METHOD

The following section presents the experimental design of the present study, by providing information on the participants, the apparatus, the experimental tasks and stimuli, and the procedure.

4.1 Participants

Three different age groups were tested in this study, namely 9-year old children (aged 9.0 to 9.8 years , *M* = 9.34, *SD* = .27), 11-year-old children (aged 11.1 to 11.7 years , *M* = 11.49, *SD* = .18), and a group of young adults between 20 and 25 years of age (*M* = 22.91; *SD* = 1.34). In each age group, four males and five females were tested. Participants were drawn from the subject pool of the Max Planck Institute for Human Development (MPI) in Berlin, either from the Centre of Lifespan Psychology or from the Centre of Educational Research. Nine participants were tested in each age group, five females and four males. Data collection took place at the MPI over a period of six months, from October 2003 to March 2004. Participants were tested individually, and they received 10 Euro per session for their participation. The study was approved by the institute's ethics committee. All participants were selected for study participation from a larger screening sample.

4.1.1 Selection of Participants

In an initial screening session, several cognitive and sensorimotor tasks were assessed. Two selection criteria for participation in the study were used: (a) difficulties in sensorimotor functioning and (b) cognitive performance.

To assess potential difficulties in sensorimotor functioning, participants filled in a questionnaire about their physical health, which listed different illnesses that might have detrimental effects on the performances required in the present study. For children, the questionnaire was completed by a parent. Participants who reported impairments of their balancing abilities caused by neurological disorders (e.g., cerebral palsy or brain tumors), acute or chronic ear infections, diabetes, recent injuries of the legs or ankle joints or who had been diagnosed with an attention-deficit-hyperactivity syndrome (ADHS) were not recruited for further testing. All participants had normal or corrected-to-normal vision and normal hearing. In addition, the body-mass index was calculated, and persons who were obese (with scores higher than the 95^{th} percentile of their respective age norms) were not asked to participate in the study. According to these criteria, three out of 23 9-year olds were excluded from further testing due to illnesses (one child with an acute ankle injury, one child with a brain tumor in early childhood, and one child whose left leg was longer than the right leg). In the 11-year olds, 19 children were screened, and three were excluded due to illnesses (one child with ADHS, one obese child with frequent ankle injuries, and one child with cerebral palsy). Of the 24 young adults that were screened, none reported any illnesses that would have led to their exclusion.

The second selection criterion was cognitive performance. Several standardized cognitive tests were used for the selection of participants to reduce the influence of differences in sample composition between the three age groups. Although clear developmental trends for most cognitive and sensorimotor tasks are reported in the literature, there is also a large variability between individuals in each age group, and the performance of an individual can differ substantially from the mean of the respective age group (Lautrey, 2003; Weinert & Hany, 2003). Because the samples in the present study were small, with only nine participants per age group, the probability of drawing samples with very heterogeneous scores was rather high. As a consequence, the correspondence of the performance scores within a sample to published age norms might differ considerably between samples.[8] If one assumes that the cognitive abilities also contribute to the pool of resources that people use to deal with the tasks, the differences in basic cognitive abilities between age groups might influence the pattern of dual-task performance. A large variation of scores within each age group and pronounced differences in sample heterogeneity between the age groups would make it harder to detect age differences in dual-task specific behaviors.

The present study therefore included three tests of basic cognitive abilities, namely the “Digit Symbol Substitution Test” measuring cognitive speed, the “Digit Span Forward and Backward” measuring memory span, and “Word Meanings” measuring knowledge of vocabulary. The tests were subtests of the German version of the “Wechsler Adult Intelligence Scale- Revised” (Tewes, 1991; Wechsler, 1981) or the “WISC III: Wechsler Intelligence Scale for Children” (Tewes, Rossmann, & Schallberger, 1999; Wechsler, 1991). Published age norms exist for these tests.

[8] For example, with only nine participants, the sample of 9-year olds might be above average in their performances in certain cognitive tests, while the young adults might be below average, simply because the random selection of participants led to that distribution by chance.

In the “Digit Symbol Substitution Test”, participants were presented a line of the digits from 1 to 9, with an abstract symbol corresponding to each of these digits printed underneath it. Participants received a sheet of paper on which several lines of digits were presented in random order, and the task was to fill in the space underneath each digit with the corresponding symbol one after the other and as quickly as possible. Children were working on that test for 120 seconds, and young adults for 90 seconds. The score was the sum of correct responses.

In the “Digit Span Forward and Backward Test”, the experimenter read out a string of numbers to the participant, at the rate of about one number per second. In the “forward” condition, the task required the participant to repeat that string of numbers in the same order as it was presented. In the “backward” condition, the numbers had to be reordered by sorting them from the last to the first number. For both the “forward” and the “backward” condition, task difficulty was increased during the testing procedure by presenting increasingly longer strings of numbers. Two strings of the same length were always presented in succession. If a participant failed to repeat both of these strings correctly, testing was ended. The score was the sum of correct strings a participant produced in both task conditions.

The “Word Meanings” test required participants to verbally explain the meaning of different nouns. Children and adults worked on different versions of the test, with 32 nouns in the adult’s version and 30 nouns in the children’s version. Nouns were presented in ascending order of difficulty, and testing was ended if a participant failed to explain three successive nouns. For adults, the score was the sum of correctly explained nouns according to the manual. For children, the manual gives the option to score zero points for incorrect or missing answers, one point when the answer is not completely correct, and two points if it is.

For these four tests, each participant’s raw score was compared to the age norms for the respective subtest from the test manuals.[9] Raw scores were translated into percentile ranks, which depict the percentage of participants in the normative sample with scores below the participant’s score. Following Cattell’s and Horn’s distinction into fluid and crystallized intelligence (Cattell, 1963; Horn, 1968), percentile scores for cognitive speed and memory span were averaged for each participant to index fluid intelligence. Participants in the range of the 50^{th} to the 90^{th} percentile for these two tests and in the 50^{th} to 100^{th} percentile for the

[9] In the children’s version of the test, norms for each three-month age bracket are published (e.g., 9-year-3-months olds to 9-year-6-months olds). For the young adults, the norms for 20- to 24-year olds and for 25- to 34-year olds were relevant in this study.

"Word Meanings" test indexing crystallized intelligence were included in the present study. Information on the distribution of percentile scores for the participants in each age group are provided in Table 1.

Table 1. *Percentile Scores of Tests Used in the Screening Procedure as a Function of Age Group*

	9-year olds	11-year olds	young adults
Percentile Score for Fluid Intelligence			
Median	73.5	56.5	73.5
Minimum	50.0	50.0	56.0
Maximum	79.0	63.0	83.0
Percentile Score for „Word Meanings“			
Median	95.0	84.0	95.0
Minimum	50.0	50.0	50.0
Maximum	99.0	99.0	98.0

Note. Percentile scores for the tests „Digit Symbol Substitution“ and „Digit Span Forward and Backward“ were averaged to obtain the percentile score for fluid intelligence.

The Mann-Whitney U Test indicated that children differ from young adults in their percentile scores for fluid intelligence (Mann-Whitney $U = 38.5$, $Z = -2.19$, $p < .05$), but not in their percentile scores for "Word Meanings" (Mann-Whitney $U = 78.0$, $Z = -.16$, $p = .876$). These differences were caused by the relatively low percentile scores for the 11-year olds compared to the other two groups. The 9-year olds were superior to the 11-year olds in their percentile scores for fluid intelligence (Mann-Whitney $U = 18.0$, $Z = -1.99$, $p < .05$), and also in their percentile scores for "Word Meanings" (Mann-Whitney $U = 18.0$, $Z = -2.0$, $p < .05$).

In sum, participants in the current study represented a positive selection of their respective age groups concerning cognitive performances, since they scored above average on standardized cognitive tests.

4.1.2 Age Differences in Intelligence Subtests

Table 2 shows means and standard deviations for the three age groups for each of the subtests. Data were transformed to compare the performances in the four tests across age groups. For the "Digit Symbol Substitution Test", scores for digits-per-seconds were calculated for each participant because children worked on the test for 120 seconds, whereas young adults worked on a 90-second version. For “Digit Span Forward and Backward”, the manuals for HAWIE-R and HAWIK required the tester to sum up the score for correctly repeated lines of digits over the forward and backward version of the test. For the purpose of comparing the age groups, I now scored the answers differently: “Digit Span Forward” and “Digit Span Backward” both received separate scores, and the score in each test depicts the longest span length in which each participant could give at least one (out of two) correct answer.[10] Scores on the test for “Word Meanings” cannot be directly compared across children and young adults, because different stimuli were used in the HAWIE-R and HAWIK. However, scores of 9-year-old and 11-year-old children can be compared, because they were based on the same version of the test.

Table 2. *Scores of Tests Used in the Screening Procedure as a Function of Age Group*

	9-year olds	11-year olds	young adults
Digit Symbol Substitution			
M	0.38	0.46	0.72
SD	0.06	0.07	0.08
Digit Span Forward			
M	5.44	6.03	7.00
SD	0.88	1.02	0.89
Digit Span Backward			
M	3.89	3.66	5.89
SD	1.05	0.50	1.45
Word Meanings			
M	30.77	38.33	25.00
SD	6.74	7.26	4.33

Note. Scores for “Digit Symbol Substitution” are expressed in items-per-second. Scores for “Digit Span Forward” and “Digit Span Backward” depict the longest span for which participants had at least one correct trial (out of two). The scores for “Word Meanings” for children and young adults cannot be compared directly, because they are based on different versions of the test.

[10] Like this, a child who repeated both examples of the four-digit strings correctly, and one example of the five-digit strings, but who fails on both examples of the six-digit strings (after which the test is terminated) receives a score of five, because this is the longest span length for which one answer was correct.

To examine age differences on the four tests, a series of one-way ANOVAs was carried out. For the items-per-second scores on the „Digit Symbol Substitution Test", an ANOVA with difference contrasts for the age group factor revealed that 11-year olds scored more items-per-second than 9-year olds ($p < .05$), and young adults scored more items-per-second than the two children's groups ($p < .01$).

In the one-way ANOVA for scores on "Digit Span Forward", 9-year olds did not differ from 11-year olds in their scores ($p = .433$), whereas the contrast comparing the two children's groups to the young adults was significant ($p < .05$). The analysis for "Digit Span Backward" resulted in the same pattern of findings: 9-year old children did not differ significantly from 11-year olds ($p = .665$), but there was a significant difference between the two children's groups and the young adults ($p < .001$).

Furthermore, 9-year-olds differed significantly from the 11-year-olds in their scores on "Word Meanings", as indicated by a t-test for independent samples with $t(16)= -2.29$, $p < .05$.

The overall picture on these cognitive measures is consistent with the developmental literature showing that cognitive speed and working memory functions improve during childhood and do not reach mature levels of performance before late adolescence or young adulthood (e.g., Case, Kurland, & Goldberg, 1982; Fry & Hale, 1996; Kail, 1991, 1992; S.-C. Li, Lindenberger, Hommel, Aschersleben, Prinz, & Baltes, 2004).

4.1.3 Age Differences in Sensorimotor Tasks

Different sensorimotor measures were assessed in the screening session. Four trials on the stable and four trials on the moving (continuous triangular movement, 3 degrees, 0.5 Hz) balance platform were conducted, each trial lasting for 30 seconds. In addition, a measure of sway tolerance has been assessed, the so-called "Functional Stability Boundary" (FSB; see also Slobounov, Moss, Slobounova, & Newell, 1998). Participants stood on the force platform and were instructed to move as far as possible in two given directions, without feeling unsafe or being afraid to fall. The knees, hips and back should be kept in a straight line. The stability boundary was assessed for (a) the anterior-posterior, (b) the lateral, (c) the diagonal right, and (d) the diagonal left direction. The calculation of the area of FSB takes the maxima in each direction into account. Table 3 presents average scores for the three age groups.

Concerning the average FSB areas, independent samples t-tests revealed that children did not differ from the young adults, $t(24) = -.106$, $p = .916$, and that 9-year olds did not differ from 11-year olds, $t(15) = 1.67$, $p = .114$. For balancing on the stable platform, young adults were superior to children, $t(25) = 3.06$, $p < .01$, but 9-year olds did not differ from 11-year olds, $t(16) = 1.63$, $p = .122$. The same pattern of findings emerged for balancing on the moving platform, with adults swaying less than children, $t(25) = 3.04$, $p < .01$, and with 9-year olds not differing from 11-year olds, $t(16) = .55$, $p = .590$.

Table 3. *Balance Measures of the Screening Session*

	Balance on Stable Platform	Balance on Moving Platform	FSB Area
9-year olds			
M	197.19	1704.79	26341.9
SD	94.39	488.72	4382.2
11-year olds			
M	132.81	1528.87	20241.7
SD	71.25	825.87	9401.6
young adults			
M	71.29	871.74	23455.2
SD	36.18	430.71	7740.7

Note. Balance performance was measured in the size of the COP area (mm^2). The first two columns depict the balance performance on the stable and on the moving platform, and the last column presents the average FSB areas. Note that the ankle-disc board was not used for balance tasks during the screening session.

4.2 Apparatus

This section describes the apparatus used for the three experimental tasks of the study: the Method-of-Loci memory task (MOL), the N-back task, and the balance task. In the single-task condition, a Power Macintosh 7100/ 80 computer was used to present the stimuli of the two cognitive tasks. The Power Macintosh was also used for the stimulus presentation of the baseline condition for balance combined with MOL, in which people were detecting animal voices in a stream of color words while balancing on the ankle-disc board. These stimuli were transmitted over infrared headphones. Additionally, the two-choice reaction-time task was assessed using the Power Macintosh. In the dual-task condition, stimuli were presented by a Pentium III personal computer on a 22’’ Sony monitor.

The Balance task consisted of balancing on a so-called “ankle-disc board” or “wobble board”, which is depicted in Figure 6. This is a device to train the body’s balance and coordination.

The board used in the current study was made from plastic. On the board, the area people stood on was a circle with a diameter of 39 cm. The lower part of the board was convex. The board was about 8 cm high, and it had a special corrugated surface to prevent the feet from sliding off.

To measure how people stabilize themselves on the board, it was placed on a 40 x 60 cm dynamic force platform (Kistler force platform 9286AA, Kistler Instrumenten AG, Winterthur, Switzerland). This measurement platform was surrounded by a second 120 x 140 cm platform, on which safety handrails were adjusted at each side at waist level. People were additionally secured by safety belts known from mountain climbing (Liberty Adjust Tape D, Mammut, Cologne, Germany).

Figure 6. The Ankle-Disc Board

These belts were fixed at an overhead frame at the balance machine, and could be adjusted to each participant's height individually. Every participant was constantly wearing the security belt during balance testing, but no participant ever stepped off the board during a trial or lost his balance while standing on the board.

Figure 7 depicts a boy standing on the ankle-disc board on the balance platform. This position is comparable to a participant's position on the platform during the measurement occasions, except that participants were facing a computer monitor at the left-hand side of the platform (a 90 degree turn to the right from the position presented in the picture). The height of the computer monitor could be adjusted individually to position the monitor at each participant's eye level. Additionally, a white sheet of paper was adjusted on the measurement platform, on which the position of the robot axis was indicated. This assured that the board was placed on the exact centre of the platform, which made the effect of platform movements comparable across participants.

Figure 7. A Child Balancing on the Ankle-Disc Board

The measurement device for balance consisted of the balance platform itself and two computers. Two monitors were used, one on the table where the experimenter sat, and the other one at the left-hand side of the platform presenting the stimuli in the dual-task situations. The measurement computer (μ-M-S Eth-RJ45, mcm Prüfsysteme GmbH, Berlin, Germany) controlled the speed and angle of platform movement, and preprocessed, amplified and integrated the primary data. The other computer (NEXOS Pentium III, 5000 MHz, PC / NT) provided either online or offline feedback of balance performance, and additionally presented the experimental stimuli to participants during dual-task trials for the MOL and N-back task. In cases in which stimuli were presented auditorily, an infrared sound transmitter was used, and participants received the stimuli over headphones.

The dynamic force platform had twelve piezoelectric sensors located in arrangements of three at each corner. It measured postural sway by recording the components of the ground reaction force on the lateral, vertical, and anterior-posterior horizontal axes, along with the three respective moment components. Signals were hardware-filtered by the measurement computer with a Butterworth third-order filter resulting in a sampling rate of 1000 Hz, removing high-frequency artifacts. The position of the body's centre-of-pressure (COP) could then be calculated for each millisecond that a participant stood on the platform. The variable of interest in the current study was the COP area, which is the area within which the COP moves in a given time. It reflects the amount of shifts in the forces applied on the platform by the body while maintaining upright stance on the ankle-disc board. The COP area represents the portion of the base of support utilized while balancing on the board. Good performance, which is achieved through small and well-coordinated body movements on the board, results in small COP areas.

For trials on the moving platform, a central robotic axis (Power Cube Rotari PR 110, mcm Prüfsysteme GmbH, Berlin, Germany) produced a triangular wave movement. When the board was placed on the moving platform, participants had to counteract the movement of the platform by shifting their weight in the opposing direction of platform movement. Platform movements were induced along the lateral plane, therefore the movement had to be counteracted by shifting one's weight to the left or right foot. Naturally, platform movements lead to larger COP areas than balancing on the stable platform, and good performance required successful and efficient stabilization of the body's equilibrium on the moving platform. Figure 8 provides an example for a small and a large COP area.

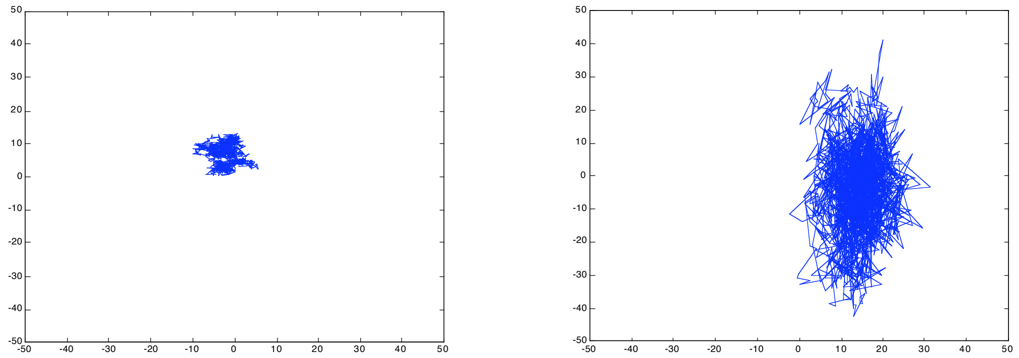

Figure 8. An Example for a Small and a Large Center-of-Pressure (COP) Area on the Force Platform

Note. The axes of each figure depict deviations in mm.

4.3 Experimental Tasks and Stimuli

Two different cognitive tasks were used in the present study, namely an episodic memory task requiring the strategic remembering of word lists (Method-of-Loci memory task), and a working memory task, in which digits had to be continuously encoded and recalled at a specific point in time (N-back task). The sensorimotor task consisted of balancing on a special device, the ankle-disc board, which was placed either on a stable or moving surface. Additionally, a two-choice reaction-time task was assessed at the beginning and end of most sessions. The following sections describe the tasks in more detail.

4.3.1 Method-of-Loci Memory Task

The MOL task required participants to memorize a list of words that they heard over headphones. Participants were instructed in the Method-of-Loci strategy (MOL), which involves the use of a highly familiar ordered sequence of locations of a mental map as a structure for encoding and retrieving new information (Bower, 1970; Kliegl, Smith, & Baltes, 1989; Lindenberger, Kliegl, & Baltes, 1992). Participants formed interactive images connecting to-be-remembered items to the locations of the mental map. Instruction encouraged the creation of images that were interactive and dynamic, and that included different sensory dimensions (e.g., smells, sounds, and movement). Exaggerations were explicitly allowed during the image-creation phase. For example, if the location cue was “bed”, and the to-be-remembered word was “frog”, participants were instructed to imagine a frog sitting on their bed. Depending on personal preferences, the image of the frog could be exaggerated on different dimensions, for example by imaging a huge frog with very bright and extraordinary colors and a rather slimy skin, croaking loudly, and jumping up and down. During recall, participants revisited the locations of their mental map, and tried to retrieve the items imagined.

For choosing the Method-of-Loci as the memory strategy for the current study, some characteristics of the task were crucial: It was assumed to be a memory strategy which is not used spontaneously, since it requires to have a pre-specified structure of loci in long-term memory available for generating the images. In previous research, it could be trained with participants of different age groups (Kliegl, Smith, & Baltes, 1989; Lindenberger, Kliegl, & Baltes, 1992; Singer, Lindenberger, & Baltes, 2003*),* and it led to substantial increases in memory performance. The use of imagery, which is a central aspect of the strategy, has been shown to improve children's recall performance considerably (Danner & Taylor, 1973; Jusczyk, Kemler, & Bubis, 1975; Yuille & Catchpole, 1973). Furthermore, imagery instruction was found to be more effective with children who are more intellectually competent (Pressley, Cariglia-Bull, Deane, & Schneider, 1987). This indicates that the selection of participants according to their cognitive performances in the present study had the additional advantage of increasing the likeliness that every participant actually profited from the strategy. Furthermore, because all participants were instructed and trained to use the same strategy, inter-individual differences in memory performance could not be explained by participants using qualitatively different strategies to handle the task (given that they follow instruction). Another important aspect of the task was that central task parameters can be manipulated to adjust it to each participant's performance level by increasing or decreasing list length and inter-stimulus interval. Longer lists with items presented in more rapid succession should lead to decreases in recall performance (Pressley & Levin, 1977).

List of locations. All participants worked on a 20-item list of locations that can usually be found in most households (e.g., bed, desk, chair, television, refrigerator…), presented in Table 4. The same list was used for children and adults.

Table 4. *List of Locations for the Method-of-Loci Task*

1) mailbox	8) desk	15) shower
2) doorbell	9) chair	16) toilet
3) front door	10) shelf	17) cooker
4) corridor	11) television	18) refrigerator
5) bed	12) sofa	19) sink unit
6) wardrobe	13) washbasin	20) kitchen cupboard
7) curtain	14) mirror	

Location cues had been recorded by a female speaker. Locations always appeared in exactly the same order, and they were explicit, such that participants received the location cue for encoding and for retrieval. Participants were instructed to mentally visit the locations of their own household.[11] Since list length varied within and between participants, each new list started at the next location of the location list. Therefore, if the last location of a given list was "television" (location 11), the first location of the following list was "sofa" (location 12). This procedure assured that each location occurred equally often for each participant.

Word materials. 40 word lists per participant were used. All the words used in the study were highly imaginable and highly concrete nouns and had been recorded by a male speaker. In the adaptive phase, the amount of words per list varied according to each individual's performance. As a maximum, children could receive up to 14 words per list, whereas young adults could receive up to 20 words per list. For the first three sessions of the study, separate lists with nonrandom pairings of locations and words were constructed. For theses lists, words were chosen from a German dictionary, with the constraint that they should be concrete nouns and very familiar to children. For the adaptive and the dual-task phase, 972 words from the word pool used by K.Z.H. Li et al. (2001) were used for the young adults. For children, fewer words were needed, but these words had to fulfill certain criteria to be suitable for children. Lindenberger and colleagues constructed a pool of 772 words that were rated by 7- to 9-year-old children ($N = 10$) as being highly familiar to them (Brehmer, Stoll, Bergner, Benoit, von Oertzen, & Lindenberger, 2004). Out of these words, 429 words were chosen for the adaptive and the dual-task phase of the present study.[12] For the construction of each list in the adaptive and dual-task phase, words were drawn randomly from the word pool without replacement.

Task presentation and recall. The pairs of locations and to-be-learned words were presented over headphones. Time frames for the auditory presentation of the locations were fixed to 1200 ms, which corresponded to the time needed to display the longest location. The

[11] In cases where certain locations did not exist in somebody's household, alternative locations were constructed with the help of the experimenter, which could be used at the respective location cue. For example, some participants reported not having a "shower" at home, but thought about their bathtub instead when the "shower" location cue appeared.

[12] Lindenberger and colleagues further selected 413 out of these 772 words that also fulfilled the criterion of generating distinguishable images. For the purposes of the present study, this criterion was considered too strict.

time frames for the presentation of to-be-learned words varied according to study phase (see Section 4.4 for details).

For the encoding phase for the MOL task, participants sat in front of the computer under single-task conditions, and they stood on the ankle-disc board under dual-task conditions. Likewise, participants recalled the items sitting in front of the computer in the single-task condition, or standing on the platform in the dual-task condition. Balance performance was not measured during recall, but only during encoding, such that participants stepped off the board and could move freely on the platform during the dual-task recall phase.

At recall, participants received the first location cue of the list right after the encoding phase. Participants were then asked to retrieve the to-be-learned item from memory. For retrieval in front of the computer, young adults typed the first three letters of each retrieved word into the according frame on the screen, whereas children named the word they remembered at that location, and the experimenter typed in the answer to avoid performance decrements caused by spelling mistakes. After an item was typed in, the program presented the next location cue. In the dual-task situations, the first three letters of the word were always typed in by the experimenter, who sat in front of a computer screen close to the balance platform. The upper time limit for the entry of each answer was always 90 seconds. If words could not be remembered by the participant, location cues were skipped without entering any answer. These instances were recorded as error by the program. At the end of each list, participants were presented a feedback screen displaying the to-be-remembered and the actually retrieved words at each location and the number of correct responses.

4.3.2 N-Back Working Memory Task

To measure working memory performance in the current study, the N-back task was used. The task requires monitoring, short-term storage, and scheduled retrieval of digits, and it is assumed to rely heavily on working memory functions. Executive control processes are also necessary for working on the task (Dobbs & Rule, 1989; Smith & Jonides, 1999).

Stimuli were digits from 1 to 9 presented successively in a white square (8 x 8 cm) on black background on the computer screen. The font of the digits was Helvetica, and digits were about 6 cm high and 4 cm wide. Digits were drawn randomly for presentation, with the constraint that at least two different digits had to be presented before a certain digit could be repeated. Furthermore, numerically adjacent digits never appeared in succession. Depending on the condition, participants were asked to read out each digit aloud as soon as it appeared (N-back 0), or they had to postpone the verbalization of the digit until the next digit was presented (N-back 1), the second-next digit (N-back 2), or the digit in the position three (N-back 3) or four (N-back 4) after that. Children were trained to work on the N-back 2 version of the task, whereas young adults worked on the N-back 4 version. Items were scored as correct if the correct digit was named in the correct time-window (beginning at the presentation of the to-be-remembered digit and ending at the presentation of the next digit). Errors occurred (a) if a wrong digit was named, which was not correct in that serial position, (b) if no digit was named at all, or (c) if a digit was named earlier than it should have been named.

Because the timing of the verbalization was crucial, two experimenters were always present to detect possible errors. One experimenter wrote down the digits that were verbalized by the participant, and the other experimenter controlled whether digits were verbalized in the correct time-windows. Following each trial, the program presented all the digits that had appeared, and these could then be compared to the digits that had been written down. If participants committed more than two errors within a trial, correct digits were only scored to the point at which the second error had occurred. This was done to assure that participants continuously worked on the task without taking breaks during a trial. Inter-stimulus intervals and number of successively presented digits varied according to the phase of the study (see Section 4.4 for details).

4.3.3 Balance Task

The typical way posturography is measured with experimental tasks is to communicate to participants to „move as little as possible“ while standing on a stable platform for a certain time (e.g., DeOreo & Wade, 1971; Figura, Cama, Capranica, Guidetti, & Pulejo, 1991). Presumably, this is not a task children encounter in everyday situations, nor does it fit their behavioral preference. This is even more problematic with rather long trial durations, which are necessary in the current study due to the trial durations of concurrent cognitive tasks and due to reliability issues (Holliday & Fernie, 1979; Le Clair & Riach, 1996). To make the task more interesting and challenging for children, I used a special balance device, the so-called “ankle-disc board”, in the present study. Standing on the board and stabilizing one’s body position requires constant small movements of different body parts, especially of the muscles and joints in the legs. The board has been designed to challenge and improve balance, stability, and proprioception (Tropp & Askling, 1988). It is often used in the rehabilitation training after injuries of the ankle joints. Less pronounced and smooth body movements represent successful stabilization of the body’s equilibrium on the board.

In the present study participants were instructed to "stand as stable as possible" on the ankle-disc board, with their arms loosely to their sides. Participants wore standard sport shoes provided by the lab during the balance trials. Feedback about the COP movement was either given online (when participants could visually follow their COP movement on the computer screen during the trial) or offline (when they received feedback about their COP movement after the trial).

Balance performance was assessed on the stable and moving platform. Under moving platform conditions, the platform tilted in the lateral plane with a triangular-shaped, continuous movement at a frequency of 0.5 Hz. The amplitude of platform movement differed according to age group, and is presented in Table 5. For each age group, two different movement conditions were used. Presentation of results focuses only on those conditions that were identical in all groups, namely the stable platform and the three degrees movement condition.

Table 5. *Degree of Platform Movement for Difficulty Manipulation of the Balance Task*

Age group	Baseline	Difficulty Level 1	Difficulty Level 2
9-year olds	stable	*1 degree*	3 degrees
11-year olds	stable	3 degrees	*5 degrees*
young adults	stable	3 degrees	*7 degrees*

Note. The degrees in italics were specific to a certain age group. Results only refer to the stable and the three degrees moving platform conditions.

In addition to balancing on the stable or on the moving platform, two different baseline conditions were used for the balance task, given that balancing had to be performed with two different cognitive tasks under dual-task conditions. The two cognitive tasks differed in (a) trial duration, and (b) factors that might influence the balance performance independent of the cognitive load of the task, like verbalization in the N-back task and the auditory perception of stimuli in the MOL task. There is empirical evidence that verbalization influences balance performance (Yardley, Gardner, Leadbetter, & Lavie, 1999). Concerning *trial duration*, MOL trials lasted for 60 seconds, and N-back trials lasted for 35 seconds. The respective single-task balance trials therefore used the same trial durations as the cognitive tasks. To control for the influence of *load-independent factors* like verbalization and auditory stimulus perception, participants were performing different secondary tasks while their "single-task" baseline performance in balancing was assessed. These tasks were designed to closely resemble the perceptual task demands of the cognitive tasks without imposing a strong cognitive load on participants.

Balance and MOL. As the reference (or "baseline" or "single-task") condition for balancing while memorizing the word lists of the MOL task, participants performed the balance task while concurrently listening to a stream of color words and animal voices. The stimuli for that task were color words (e.g., orange, blue, yellow; recorded by a male and a female speaker) or animal voices, which could be a dog barking, a cow mooing, or a goat bleating. The number of stimuli presented in the animal voices task and the inter-stimulus intervals (ISIs) between successive stimuli corresponded to the task parameters a participant encountered in a MOL list of that study phase. Participants were instructed to focus attention on the animal voices. Each animal voice could be presented once within a trial, such that no animal, one animal, two animals, or all three animals could appear. Participants were asked to report directly after the end of the trial which animals they had heard, irrespective of the order in which they had appeared. In the dual-task condition, participants were balancing on the ankle-disc board while they concurrently encoded the MOL stimuli. Trials for balance in combination with the MOL task or the animal voices task always lasted for 60 seconds.

Balance and N-back. The reference (or "baseline" or "single-task") condition for balance and N-back required participants to simply read out aloud digits presented on the computer screen during the balance trial. The number of (and ISI between) successive digits depended on each participant's task parameters for the N-back task in that phase of the study. In the dual-task condition, participants performed the N-back 2 (children) or N-back 4 (adults) task while balancing. Trials for balance in combination with the N-back task always lasted for 35 seconds.

4.3.4 Two-Choice Reaction-Time Task

Since up to six different word lists of the MOL task were administered in each session of the study, proactive interference might have decreased the recall performance of the last lists in a session. Proactive interference refers to the phenomenon that images from previous lists interfere with the current words to-be-learned (Kliegl & Lindenberger, 1993). To be able to distinguish the effect of proactive interference from mere tiredness, a computerized two-choice reaction time task was administered at the beginning and the end of sessions 2 to 8, with two 30-second trials each. Participants were sitting in front of the computer screen and were holding a button box in each hand, one with a yellow and the other one with a red button on it. They were instructed to react as quickly and accurately as possible to two easily distinguishable visual signals (a big red spot versus a big yellow spot appearing on the screen in a central location) by pressing the respective button with their thumb. The spots appeared at fixed time intervals with an ISI of 1200 milliseconds. There were 24 targets in each trial. Online feedback about reaction accuracy was given by a high-frequency tone for each correct reaction and low-frequency tone for each error (omissions and intrusions). The pairing of keys and colors was counterbalanced across participants. The program recorded reaction times to each target and errors.

4.4 Procedure

Data were collected in the course of nine sessions lasting for about one hour each. Participants were tested individually, with two experimenters present at all times. Table 6 presents an overview on the study design, and more detailed information is provided in Table D.1 in Appendix D.

Table 6. *Overview of the Study Design*

	Task		
Session of the Study	MOL Task	N-Back Task	Balance Task
Session 1	Pretest	Training	Training on Stable Platform
Session 2	Instruction	Training	Training on Stable Platform
Session 3	Training	Training	Training on Stable and Moving Platform
Sessions 4 and 5	Adaptive Training (Adjustment of Task Difficulty)	Adaptive Training (Adjustment of Task Difficulty)	Training on Stable and Moving Platform with Secondary Tasks
Sessions 6, 7, and 8	Dual-Task Assessment with Different Difficulty Levels of the Balance Task		
Session 9	Dual Task Assessment with Differential-Emphasis Instruction		

The first three sessions were used to instruct and train participants in the three component tasks (MOL, N-back, and balancing) under single-task conditions. In the adaptive phase, consisting of Sessions four and five, task difficulty of the cognitive tasks was manipulated, with the aim to adjust each individual's performance to the level of about 80 % correct. The assessment of dual-task performance took place in Sessions six to eight, in which N-back performance or MOL performance was measured while participants were concurrently balancing on the ankle-disc board. In the last session, dual-task performance was measured under differential-emphasis instructions, when participants were told to focus more strongly on one task domain than on the other.

4.4.1 Instruction and Practice

MOL. In the first session, participants were asked to memorize lists of paired loci and words without having been instructed in the Method-of-Loci memory strategy yet (MOL pretest). MOL instruction took place in the second session, and the third session was used to practice the MOL task. The ISI was fixed for the first three sessions, except for the instruction lists in the second session, in which the presentation of the stimulus was self-paced, to allow for enough time to discuss the creation of suitable mental images with the experimenter. Table A.1 in Appendix A lists the fixed ISIs and the list lengths for the first three sessions of the study. Note that the ISI was defined as the time window in which the word was presented (excluding the fixed time window for the location).

N-back. In the first session, when the task was introduced, children also worked on the N-back 1 version and young adults on the N-back 1, 2, and 3 version for several trials to practice the task. ISIs and number of successively presented digits were fixed in the first three sessions of the study and are presented in Table B.1 in Appendix B. Within the first five sessions of the study, trials were repeated if participants had named less than three (children) or five (adults) correct digits.

Balance. Online feedback was only used in the first session of the study to familiarize participants with the platform and measurement procedure. Balance performance was assessed initially without any secondary task to let participants practice balancing on the ankle-disc board. After several practice trials on the stable platform in the first two sessions, platform movements were introduced to participants in the third session. Table D.1 in Appendix D presents the details.

4.4.2 Adaptive Phase

In the adaptive phase of the study, task difficulty was manipulated individually for the two cognitive tasks. List length and ISI varied concurrently, and high performance levels in previous trials (MOL) or blocks (N-back) led to more items and shorter ISIs in the following trial or block, with the aim of calibrating each participant's performance at 80 % correct. The algorithms that generated the list length and ISI for the following trial took the difference between current performance level and 80 % target level into account, such that larger adjustments were made when the performance differed considerably from the target level than when it was rather close to the target level. At the end of the adaptive phase, each participant's data was used to select an individually adjusted task parameter setting (as a fixed combination of a certain list length and ISI) that then remained constant for the following sessions of the study. For each individual, the most difficult parameter setting (longest list length, shortest ISI) in which the participant had scored at least 70 % correct at least twice within the adaptive phase was selected. This procedure allowed to take the performance of the entire adaptive phase into account, rather than performance on the last trial only.

MOL. Table A.2 in Appendix A shows the possible list lengths and the according ISIs for the last six sessions of the study. The list length with which participants started the adaptive phase depended on age group. Children started the adaptive phase with six words and adults with eight words. Tables A.3 and A.4 in Appendix A present details of the algorithm used. Table A.3 summarizes the 80 % correct performance for the different list lengths of the study, and Table A.4 presents the algorithm used to calculate the list length of the following trial during the adaptive phase. If a participant's recall performance was more than two items above or below the 80 % level, the algorithm increased or decreased task difficulty of the following trial by two units (steps). If the participant was already close to the 80 % correct recall performance, task difficulty was adjusted by one unit only.

N-back. Task difficulty was adjusted after each block of trials in the N-back task, taking the average performance of three successive trials into account. Table B.2 in Appendix B provides the possible combinations of list lengths and ISIs for the N-back task in the adaptive and the dual-task phase of the study, and Table B.3 in Appendix B presents the algorithm used for manipulating task difficulty.

Balance. From Session four on, balance was always accompanied by one of the secondary tasks. During the adaptive phase, participants practiced to perform the balance task while concurrently reading numbers or listening to animal voices. The platform was either stable or moving. Table D.1 in Appendix D provides the order in which the different balance trials were assessed.

4.4.3 Dual-Task Assessment

Under the dual-task conditions of Sessions six to eight, participants were instructed to stand as stable as possible and to score as high as possible on the concurrent cognitive task. Each of the dual-task sessions was conducted with a specific difficulty level of the balance task.

4.4.4 Differential-Emphasis Instruction

In the last session of the study, in which dual-task performance under differential-emphasis instructions was assessed, the N-back task was performed concurrently while balancing on the stable platform. There were three different instruction conditions, (a) “focus more on the N-back task”, (b) “focus more on the balance task”, or (c) “focus on both tasks equally”. A reinforcement scheme was used to motivate participants to actually follow these differential-emphasis instructions. Each individual’s dual-task performance of the previous session with balance on the stable platform and N-back was used as the reference standard. The experimenter calculated a personal goal-criterion in each task domain (balance and N-back), which required an improvement of about 20 % for every task. If participants reached that criterion, they received credit points. More credit points could be gained in the task domain that people were instructed to focus on than in the alternative task domain. The points were summed up at the end of the session and could be traded into different kinds of sweets. Table C.1 in Appendix C presents the trade-off scheme for the credit points under the three emphasis instructions.

4.4.5 Order of Tasks and Counterbalancing

Table D.1 in Appendix D includes information on the order of experimental tasks within each session. Single-task performance for every task was also assessed in each of the dual-task sessions, in order to establish a reliable baseline of single-task performance over the course of the dual-task phase. To control for the influence of within-session practice effects or the possibility of increasing tiredness in the course of the testing session, all dual-task sessions of the study followed an ABBA-scheme concerning the administration of single- and dual-task trials. According to that scheme, single-task performance was assessed at the beginning and end of each session, and dual-task performance was assessed in between.

For each dual-task session, one specific balance-difficulty condition was used, which remained constant for the entire session. Table 7 presents the order of movement difficulties for the dual-task phase (Sessions six to eight) depending on ID.

Table 7. *Counterbalancing for the Balance Difficulty of Dual-Task Sessions*

IDs	Session 6	Session 7	Session 8
....1 ....4 7	stable	movement difficulty level 1	movement difficulty level 2
....2 ....5 8	movement difficulty level 1	movement difficulty level 2	stable
....3 ....6 9	movement difficulty level 2	stable	movement difficulty level 1

Note. IDs used in the study always ended in the digits 1 to 9 within each age group, and participants were randomly assigned to IDs.

In Session nine, the order of the differential-emphasis conditions “focus on N-back” and “focus on balance” was counterbalanced across participants, such that participants with IDs ending in even numbers were working on the “focus on balance” condition first, while participants with odd numbers were working on the “focus on N-back” condition first. Performance feedback was given after every trial in each task condition.