

Article

Not peer-reviewed version

Fine Tuning of an Advanced Planner for Cognitive Training of Older Adults

[Mauro Gaspari](#)^{*}, [Giovanna Mioni](#), Dario Signorello, [Franca Stablum](#), Sara Zuppiroli

Posted Date: 5 July 2024

doi: 10.20944/preprints202407.0458.v1

Keywords: Cognitive Training; Planning; Executive Functions; User Studies; Heuristic Search Planner



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Fine Tuning of an Advanced Planner for Cognitive Training of Older Adults

Mauro Gaspari ^{1,*}, Giovanna Mioni ², Dario Signorello ², Franca Stablum ^{2,3}
and Sara Zuppiroli ¹

¹ Department of Computer Science and Engineering, University of Bologna; mauro.gaspari@unibo.it

² Department of General Psychology, University of Padova; franca.stablum@unipd.it

³ Human Inspired Technology Research Centre, University of Padova

* Correspondence: mauro.gaspari@unibo.it; Tel.: +39-051-209487

Abstract: Developing effective cognitive training tools for older adults, specifically addressing executive functions such as planning, is a challenging task. It is of paramount importance to ensure the implementation of engaging activities, which must be tailored to the specific needs and expectations of older adults. Furthermore, it is essential to provide the appropriate level of complexity for the planning task. A human-centred approach was employed to address the issues identified in the design of the tool. Two pilot studies were conducted with older adults to fine-tune the training task and optimise its suitability for them. This also led to an enhancement of the underlying planning engine, transitioning from simple fast-forward planner (PDDL4J) to an advanced heuristic search planner (ENHSP). The results show that conducting user studies enabled the development of a cognitive training system that gradually increases the proposed difficulty levels of the planning task while maintaining usability and satisfaction among older adults. This highlights the importance of conducting user studies when implementing cognitive training tools for older adults.

Keywords: cognitive training; planning; executive functions; user studies; heuristic search planner

1. Introduction

The world is currently undergoing a profound demographic shift, moving from a population structure in which the majority of individuals were relatively young to one in which a significant proportion of people are over the age of 65. According to data from the World Population Prospects: the 2022 Revision [1], by 2050, one in six people in the world will be over the age of 65 (16%), with this figure rising to one in four for those living in Europe and Northern America. In 2018, for the first time in history, persons aged 65 or above outnumbered children under five years of age globally. The number of persons aged 80 years or over is projected to triple, from 143 million in 2019 to 426 million in 2050 [2]. This change presents both a challenge and an opportunity for the design of intelligent technology for aging [3]. Cognitive health is a significant factor in determining the functional ability of older adults [4–6], and it is of paramount importance for maintaining autonomy.

The development of cognitive training programmes is becoming a priority for reducing the impact of ageing on quality of life. One of the key aspects of successful ageing is the ability to solve everyday problems encountered in daily life. Any task that requires planning, organisation, memorisation, time management, and flexible thinking is particularly challenging for older adults. Retirement and withdrawal from productive activities often leads older people to limit activities and refrain from using problem-solving skills as previously done. Consequently, individuals may encounter greater difficulty in finding a successful solution to a problem as they age. Previous studies have indicated that cognitive training, even when initiated later, can have positive benefits, with reduced rates of cognitive decline and lower incidence of dementia [7,8].

Based on this evidence, brain games, initially available in a paper-and-pen format, have been designed and implemented on computers to train problem-solving abilities [9]. To be effective, the training tasks should have high ecological validity (training participants to perform activities typical of everyday life), be easily usable, and be sufficiently engaging. This can minimise the number of

individuals who abandon the training, thereby increasing the number of subjects who can benefit from the training. For example, an engaging scenario can be designed to simulate a visit to a historic city with several constraints and goals to achieve within a limited timeframe.

In this context, the use of AI technology, such as automated planning, has been shown to be beneficial in crafting realistic and engaging scenarios with human-centered features [10,11]. For instance, it can be employed to generate and assess a range of scenarios in which older adults can exercise their problem-solving abilities while taking into account a set of goals and constraints.

From a technological perspective, the design of a cognitive training system based on planning that supports the aforementioned features poses several challenges:

- the design of engaging problems for older adults;
- the accurate determination of the appropriate difficulty levels of the exercises;
- the design of a mechanism to adapt the difficulty of exercises to subjects throughout the training.

To address these challenges, it is essential that the design and implementation of cognitive training tasks involve older adults. This should be done in a way that builds a solution specifically conceived for them, following a participatory design approach.

This paper presents the approach taken to address the challenges identified in the SWIFT (Shared, Web-based, Intelligent Flexible Thinking Training) project. The project aims to develop a framework to support problem-solving training for older adults. The SWIFT framework consists of a platform that provides a set of training tasks, a user interface for older adults, and one for administrators, enabling them to configure and monitor training sessions.

The proposed task requires users to plan a two-day vacation in a European city (Rome). This involves organising virtual train and hotel reservations, as well as undertaking various activities (e.g., visiting specific locations and attending particular events). This scenario is encoded as a planning problem, allowing for the creation of different instances of the problem, each featuring unique goals and constraints.

In the course of our development process, we employed a participatory design approach to address the aforementioned challenges. Following the development of a first prototype [12], a Focus Group Study was conducted [13] to identify and address the requirements of the training task. The resulting system underwent further refinement through two primary pilot studies. These user studies enabled us to fine-tune the difficulty levels of the planning task to adapt to older adults throughout the training while maintaining high usability and ecological appearance standards. Furthermore, preliminary evaluation of the effectiveness of the proposed training yielded intriguing results. Consequently, the crucial role of user studies in the development of complex cognitive training tools is emphasised.

2. Methods

The design of ecological training tasks for executive functions necessitates a coordinated multidisciplinary research effort. On one hand, complex technical solutions are required, such as adapting to the difficulty of performing exercises or exploiting automated planning techniques for the exercises. On the other hand, the supervision of cognitive psychologists is of paramount importance. Testing with subjects is essential for tuning tasks before delivery. It is also essential to consider issues such as personalisation and adaptability when working with older adults [14]. Indeed, the reduced plasticity in ageing necessitates a higher level of customisation and adaptability.

The cyclic development process is depicted in Figure 1. The development of tasks is divided into six macro phases based on a cyclic structure. The sequence of the phases is not fixed; movement between them is possible in both directions. The outcome of each phase determines which phase has to be performed next. A working version of the software is produced during the first step, so experimentation can start early in the software life cycle. Each subsequent release of the task incorporates new functions or rectifies any deficiencies present in the previous release.



Figure 1. Graphical representation of the cyclic development process.

The identification of the task commenced with an initial prototype of a cognitive training task, named Weekend in Rome referred to as Version 0.0 (V0.0), which required users to plan a two-day vacation in Rome [12]. This prototype was subsequently enhanced (V1.0) through the involvement of older adults in focus groups [13], with the objective of addressing fundamental requirements. Subsequent to this, two pilot studies were conducted, with the objective of fine-tuning the task and its difficulty levels. A usability study (A) was conducted to assess user satisfaction, which enabled the prototype to be refined (V2.0). An evaluation study (B) was then carried out to assess the system's effectiveness and gather preliminary results. The following sections present the details of the training task and the two pilot studies.

2.1. The Weekend in Rome Task

In the Weekend in Rome task, users have to organise virtual train and hotel reservations and to complete various activities, such as visiting specific locations and attending particular events. To execute these tasks, users have to navigate a map where the goals are those typically encountered in real-life planning of trips (e.g., making reservations, checking bus schedules, and noting opening hours of specific locations). This scenario is encoded as a planning problem using PDDL (Planning Domain Definition Language) [15]. This approach enables the generation of numerous instances of the problem, each featuring different goals and constraints. This is possible because the planner can be used to assess the feasibility of each instance.

The system proposes three main stages of difficulty, designated as easy, medium, and difficult. Each stage comprises at least three distinct instances of the problem, each of which must be solved twice in order to advance to the subsequent level. The easy stage is characterised by a map in which each point can only be reached on foot, there are eight Points Of Interest (henceforth referred to as POIs) placed on the map, and the user is required to solve from a minimum of three goals to a maximum of five. In the medium stage, a map is presented where some connections are possible only with the use of buses. Buses operate on a scheduled basis, with specific times of operation indicated on the map. Additionally, the map includes a second railway station, from which users can embark or disembark. An illustrative example of this stage is presented in Figure 2. In the medium stage, users are required to achieve a number of goals, ranging from a minimum of six to a maximum of eight. In the difficult stage, a new POI is added to the map, and users are asked to achieve a minimum of seven

and a maximum of ten goals. For each stage, three instances of the task are provided with increasing difficulty levels. In order to complete their training, users must finish all difficulty stages, consisting of nine tasks, planning their journey by achieving at least 80% of the goals in each task.

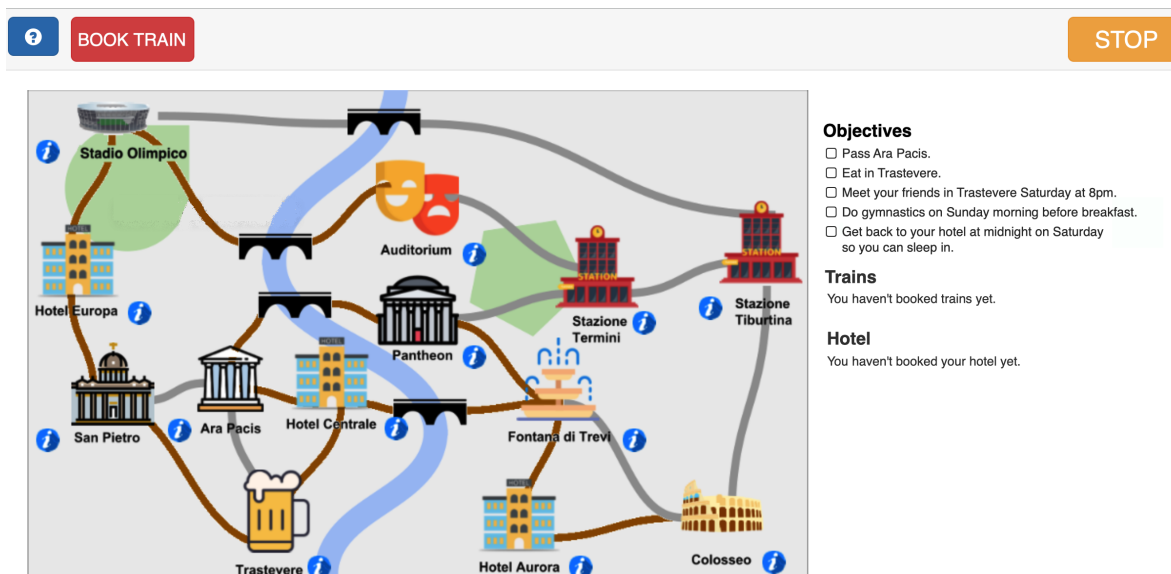


Figure 2. Example of the medium stage. The user interface for the Weekend in Rome task is designed to be simple and not confusing for older users. It presents a map of the city, the goals, and some buttons that appear or disappear dynamically to facilitate its use. The user can navigate the map using the mouse by clicking on POI adjacent to the one where they are.

Three types of goals can be achieved: a simple passage from a POI (e.g. visited Pantheon); a visit at a POI, which must take place within the opening hours of the attraction (e.g. done-activity Colosseum); a visit at a POI at a given time, for doing a specific activity (e.g. done-activity-timed Olympic Stadium at 18). Although, the developed exercise is specific to Rome, the structure can be implemented for any European city.

2.2. Exploiting Automated Planning

Versions 0.0 and 1.0 of the Weekend in Rome prototype were based on an automatic planner, PDDL4J (Planning Domain Description Library for Java) [16]. The planning domain is described using PDDL 1.2, which also allows the specification of several problems to be solved dynamically.

The planning domain encodes a set of PDDL rules, which encompasses all possible actions and interactions with the user. These include travel (e.g., walking, bus, and train), activities to be carried out at a POI (e.g., visiting, visiting at a certain time), sleeping and having breakfast in a booked hotel, and exercising. A planning problem, in accordance with the specified difficulty stage, incorporates the specific activities, bus and train timetables, connections between the various points on the map, and the goals to be achieved.

The planner is employed in different phases of the training process, namely for the generation of new solvable exercises and for the evaluation of solutions. The interaction between the user and the planner is depicted in Figure 3. The Trip Generator is activated when the user is required to undertake a new instance of the task (1). It takes as input the user profile and the level of difficulty of the new task (2), and generates a new problem instance by extracting the goals to be achieved by the user from a set of possible goals randomly (3a). Successively, the Trip Generator calls the Planner to find a plan that solves the new instance of the problem (4). If the Planner fails, steps (3a) and (4) are repeated, and other goals are selected until a solvable scenario is created (3b). Once a solvable scenario is generated, the user can start to execute the task (4). At the conclusion of the exercise, upon the user's completion of their visits by taking the return train (5), another component is initiated, the Trip Evaluator (6).

The Trip Evaluator quantifies the number of goals attained, assigning a percentage rating to the user, ranging from 0 to 100. A difficulty level is deemed to have been successfully completed when the obtained percentage is at least 80%. Consequently, it is possible to pass the instance of the problem even if the plan has not been fully executed. Furthermore, the Trip Evaluator provides feedback to the user on the plan implemented (7). For example, "Congratulations! You completed this exercise without any errors" when a difficult exercise is passed, or "This exercise was much more difficult than the previous one. Try to keep track of bus schedules" in case of failure.

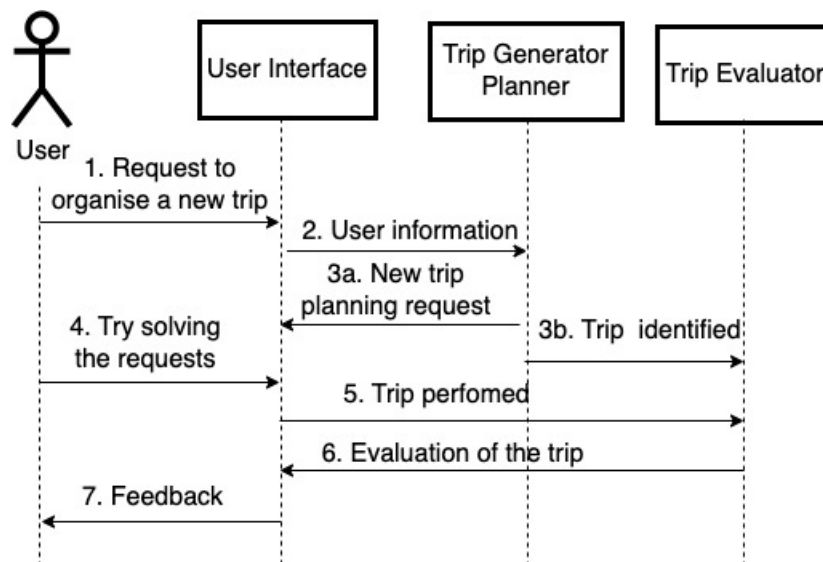


Figure 3. The interaction between the user and the planner.

2.3. Pilot Study A: Testing Usability and Satisfaction

A pilot study (A) was conducted with a group of healthy young and older adults to assess the usability of the Weekend in Rome (V1.0) task. The aims of the pilot study were the following:

1. to provide a preliminary validation of the usability of the training task on healthy older adults;
2. to identify the specific requests and needs of the older adults when performing the task;
3. to identify processing characteristics specific to older adults by examining differences in performance between older and younger adults;
4. to test the difficulty stages proposed by the system;
5. to collect all the relevant suggestions proposed by the participants.

A total of 22 participants were recruited for the study, comprising 11 young adults (aged 18-26 years) and 11 older adults (aged 62-83 years). The young adult group, comprising three males and eight females, had an average age of 23.64 (SD=1.12) and an average of 17.27 years of education (SD=1.01). The older adult group, comprising six males and five females, had an average age of 68.73 (SD=7.55) and an average of 11.45 years of education (SD=1.96).

All participants completed a series of 40-minute training sessions until they had experienced all the difficulty levels. Prior to the commencement of the study, all participants signed the Research Informed Consent Form and received written instructions for accessing and utilising the training tool. The young adults completed the online sessions independently, while the older adults were supervised until they demonstrated satisfactory compliance with the tool. All participants were instructed to contact the experimenter should they require further information or clarification. The sessions were monitored using the remote-control facilities provided by the system. At the conclusion of the sessions, all participants completed a usability questionnaire.

2.3.1. Results of the Pilot Study A

The results of the usability questionnaire indicated that all the older participants were able to easily access the online system. Furthermore, the majority of them (9/11) reported no difficulties in understanding and performing the task. The responses were based on a Likert scale (1= not at all; 4= a lot). The majority of the older participants indicated that good planning abilities ($M=3.64$; $SD=0.50$) and computer experience ($M=3$; $SD=0.63$) were crucial for completing the task. With regard the gradual increase in difficulty, the responses indicated a limited satisfaction ($M= 2.45$; $SD=0.93$). Furthermore, the item related to the ecological quality of the task indicated a need for improvement ($M= 2.73$; $SD=0.90$). Older adults identified the high involvement, the possibility to improve their problem-solving and planning abilities, the engaging and challenging task format, and the new technological approach as the main strengths of the task. Older participants offered a number of suggestions for improvement. These included making the task goals visible on the map at all times, streamlining the train booking process, adding new places to visit, changing the colour of the streets in the map to enhance visibility, adding new actions related to a real journey (i.e., the introduction of a budget for the trip to cover expenses of hotels, trains and buses). All participants successfully completed the task.

A t-test was conducted on the critical dependent variables with Group as the between-subject factor (young vs older). The following performance variables were evaluated: number of not achieved goals, execution time (minutes), number of clicks on reservations, number of clicks on goals. See Table 1 for the results. A significant difference ($p < 0.05$) was found between the two groups in the number of not achieved goals, execution time, and number of clicks on reservations. The older adults showed a higher number of not achieved goals, a longer execution time, and a higher number of clicks on reservation, indicating that they check the train and hotel reservations more often.

To test the difficulty stage proposed by the Trip Generator, a repeated measures ANOVA was carried out on execution time. The between-subject factor was Group (young and older adults), while the within-subject factor was difficulty stage (easy, medium, and difficult). The simple effects of group [$F(1, 20) = 20$, $p < .001$] and difficulty stage [$F(2, 40) = 17.43$, $p < .001$] were significant. It is noteworthy that the interaction between group and difficulty stage was significant [$F(2, 40) = 5$, $p = .012$] (see Figure 4). Older adults were slower than younger adults for all the difficulty stages. Furthermore, older adults were slower in the medium and difficult stages relative to the easy stage. Interestingly, no significant difference emerged between the medium and the difficult stages for both young and older participants.

Table 1. Means and Standard Deviations for the critical variables of the task.

Variables	Young Group M(SD)	Older Adults M(SD)	p
Number of sessions	2.09 (0.53)	4.27 (1.10)	< .001
Not achieved goals	8.73 (5.85)	13.82 (6.08)	.05
Execution time	79.95 (31.94)	176.52 (66.66)	< .001
Clicks on reservations	18.73 (11.81)	32.55 (14.67)	.02
Clicks on goals	86.91(22.88)	110.7(40.3)	0.12

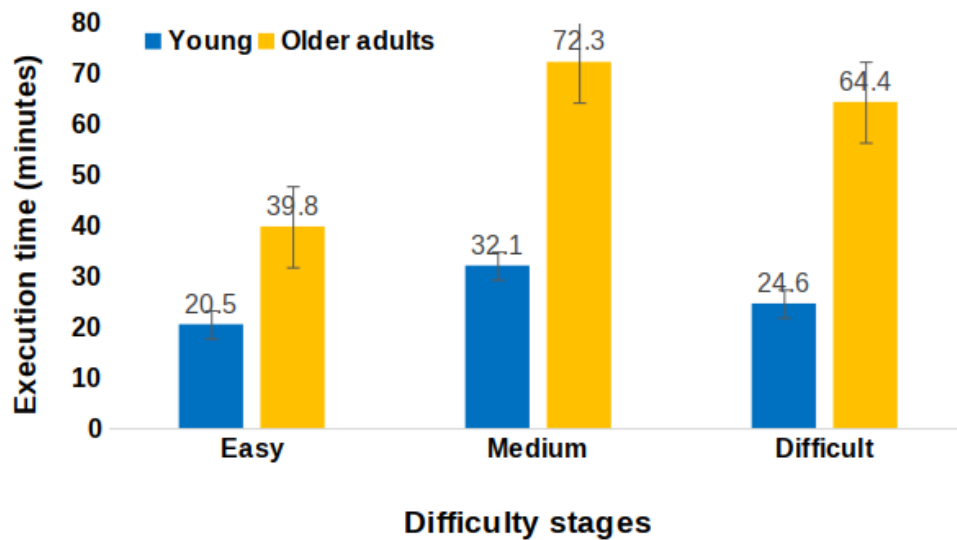


Figure 4. Means and standard deviations of the execution time (in minutes) in the easy, medium and difficult stages for young and older adults.

2.4. The revised version of the Weekend in Rome task

The results of the pilot study A permitted the identification of several areas for improvement in the training task, particularly in relation to the utilisation of automated planning.

With regard to the user interface, the pilot study A confirmed that it was well designed and not confusing for older participants. Nevertheless, in response to the suggestions of the participants, several modifications were implemented. These included improvements to the map visibility, the display of task goals, and the train reservation procedures. In addition to conventional trains, the latter now include high-speed trains. To this end, a new panel was added in the interface to the right of the map presented in Figure 2, which eliminates the need for repeated clicks on the reservations and goals buttons. Another suggestion was to make the task more similar to a real journey. To this end, new locations to visit were incorporated, short videos were created for specific POIs to present general information and their history, and new actions to accomplish were added.

However, several comments were not related to simple updates of the user interface, but rather had a strong implication on the system architecture. For example, the introduction of a limited budget for the trip to cover expenses of hotels, trains and buses. This new feature affected both the planner and the user interface, the latter with the introduction of a spreadsheet for expenses and simulation of credit card payments.

Considering the planner, a key objective was to enhance the progression of difficulty stages, implementing nine increasing difficulty levels (three for each stage).

In Version 1.0 of Weekend in Rome, the progression of difficulty levels was based on increasing the number of goals tied to specific times and introducing, from easy to intermediate levels, the presence of buses that allowed travel between points on the map only at certain times. These constraints led to a reduction in the number of possible plans for solving the problem and an increase in the number of steps required to solve the plan. The planning problem was encoded in PDDL 1.2, where time and movements were managed by predicates, and the actions that the user could perform on the map were described by actions in the domain.

While the introduction of buses from easy to medium stages allowed for an adequate increase in difficulty, we did not observe the same effectiveness in simply introducing a greater number of goals during pilot study A. Upon analysis of the results, it became evident that an increase in the difficulty stage did not always correspond to an increase in user difficulty. More precisely, it was observed that moving from the medium to the difficult stage, merely increasing the difficulty of the activities with

more stringent time constraints and/or adding new goals did not necessarily result in an increase in the real and perceived difficulty: participants completed the training for the medium and the difficult stages in the same execution time. It became evident that the progression of difficulty stages in the final part of the exercise was not as steep as it could have been. In essence, the time required to solve these tasks and the number of attempts to pass them decreased on average, whereas an increase in both was expected and more appropriate for cognitive training. It was determined that this effect was caused by the rule used for passing to the next difficulty level, which applies when an 80% performance is obtained. The issue was that the minimum number of goals required to achieve the threshold was not changing in accordance with the progression of difficulty stages.

To address this issue, the rule for advancing to the next level was updated, requiring participants to achieve all the proposed goals and execute the plan without any errors. Moreover, additional goals and constraints were introduced to further reduce the number of possible plans to reach the correct solution. These included requirements to minimise the expense of the trip or the time spent in the city at the difficult stage. The introduction of the budget variable was intended not only to enhance the ecological value of the game but also to increase the difficulty of the exercises. At the difficult stage, three minimisation objectives were identified: one on time, one on costs, and one encompassing both time and travel costs.

Since these features were not supported by the planner used in the versions 0.0 and 0.1 (PDDL4J), which only supports PDDL 1.2, it became necessary to rewrite the domain using PDDL 2.1, introducing functions for the representation of time and budget. Consequently, an upgrade was implemented with the Expressive Numeric Heuristic Search Planner (ENHSP) [17], which supports fluents and plan metrics as required by PDDL 2.1. Figure 5 provides an illustration of the transition from PDDL 1.2 to 2.1. It presents the encoding of the action "travel-by-train," which implements the train trip to Rome, utilising budget and time as fluents. The definition of the planning domain was improved by employing a more expressive language. Indeed, several predicates were required to implement the progression of time in PDDL 1.2. These were used to represent that two time instants are consecutive and to state that they are not in the future anymore when the action is executed. In contrast in PDDL 2.1, increasing the time variable was sufficient.

PDDL 1.2	PDDL 2.1
<pre> [action travel-by-train :parameters (?src ?dst - place ?departure-time ?hour1 ?arrival-time - time) :precondition (and (at ?src ?departure-time) (consecutive ?departure-time ?hour1) (consecutive ?hour1 ?arrival-time) (train-path ?src ?dst) (future ?arrival-time) (future ?hour1) (train-scheduled ?src ?dst ?departure-time) (in ?src)) :effect (and (not (future ?departure-time)) (not (future ?hour1)) (not (future ?arrival-time)) (not (at ?src ?departure-time)) (at ?dst ?arrival-time) (visited ?dst) (not (in ?src)) (in ?dst)) </pre>	<pre> [action travel-by-train :parameters (?t - person ?tr - train ?src ?dst - place) :precondition (and (at ?t ?src) (= (time) (train-scheduled ?tr ?src ?dst)) (>= (total-initial-budget ?t) (budget-required-train ?tr ?src ?dst))) :effect (and (not (at ?t ?src)) (at ?t ?dst) (visited ?t ?dst) (increase (time) (time-required-train ?tr ?src ?dst)) (decrease (total-initial-budget ?t) (budget-required-train ?tr ?src ?dst))) </pre>

Figure 5. The implementation of the "travel-by-train" action, comparing PDDL 1.2 with PDDL 2.1.

2.5. Pilot Study B: Testing Difficulty Stages, Usability and Effectiveness

A new version of the Weekend in Rome task (V2.0), including all the illustrated changes, was delivered. Subsequently, a pilot study B was designed to test the progression of the updated difficulty stages, the usability of the improved version, and to gather preliminary effectiveness results. The specific objectives of the study were as follows:

This second pilot study (B) involved a cohort of healthy older adults only.

The objectives of this pilot study were as follows:

1. To assess the actual rise in difficulty compared to the previous version.
2. To validate the usability of the system, including the collection of suggestions and the assessment of participant satisfaction.
3. To assess the improvement in ecological appearance.
4. To test whether there are improvements in the trained cognitive ability, specifically planning and problem-solving skills.
5. To assess the trained cognitive abilities three months after the training.

The study comprised a sample of 22 participants (aged 67-81 years), divided into an experimental group and a control group. The selection criteria for participants in both groups were as follows: individuals aged 65 and above, with no cognitive and/or psychiatric disorders. The experimental group, comprising seven males and four females, had an average age of 72.72 years (SD=4.90) and an average of 12.36 years of education (SD=5.20). The control group, comprising six males and five females, had an average age of 70.18 (SD=4.35) and an average of 13.65 years of education (SD=3.75). Prior to the commencement of the study, all participants signed the Research Informed Consent Form. The experimental group received written instruction for accessing and utilising the training tool. The training phase was delivered exclusively to the experimental group and comprised eight training sessions (two per week), each lasting 40 minutes, using the Weekend in Rome task (V2.0). The training sessions were monitored, with 10 out of 11 participants being observed in person and one remotely using the facilities provided by the SWIFT platform.

The participants were assessed at three distinct time points: T1, the test phase, at the beginning of the study, to establish baseline performance; T2, the re-test phase, soon after the training, five weeks afterwards the test phase; and T3, the follow-up phase, three months after the re-test phase, exclusively among the experimental group. The assessments were administered to all participants at the Department of General Psychology (Padova). The following tests were administered: the Behavioural Assessment of the Dysexecutive Syndrome (BADS) [18,19] and the Everyday Problem Test (EPT) [20,21].

The BADS is a battery for the assessment of executive functions, comprising six subtests. The Rule Shift Cards Test assesses the ability to inhibit a previously learned response mode. This test is designed to assess cognitive flexibility. The Action Program Test assesses the ability to develop an action plan to solve a problem. The Key Search Test assesses the ability to plan actions and monitor one's performance. The Temporal Judgment Test assesses the ability to predict and estimate time. The Zoo Map Test assesses the subject's ability to plan and minimise errors through self-monitoring. The Modified Six Elements Test assesses the subject's organisational ability, shifting ability and behavioural control. Each test is associated with a specific scoring method and is calibrated to establish cut-offs based on the age of the participant and the execution time. The EPT is a test of everyday problem-solving, with a focus on performance accuracy. It presents real-world problems covering all seven instrumental activities of daily living domains (household management, transportation, meal preparation and nutrition, financial management, health, shopping, and telephone skills). The abbreviated (14-item) and parallel (14-item) versions of the Italian adaptation of the test were employed. One point is awarded for a correct answer, while zero points are given for an incorrect response. Subsequently, the scores are adjusted according to age and educational level cut-offs.

Furthermore, usability and satisfaction questionnaires were administered at the conclusion of each training sessions, as was the case in Study A. Additionally, participants were invited to provide

suggestions regarding potential modifications to enhance the training task and the SWIFT platform user interface through interviews.

3. Results

Although the duration of the proposed training was limited to eight sessions, the results of this study yielded several insights. The primary findings pertain to the significant enhancement in the degree of difficulty observed with respect to the initial Weekend in Rome prototype. To this end, we compared the data obtained from the two groups of older adults who underwent the eight sessions training in the pilot studies A and B. The results are summarised in Tables 2 and 3.

Table 2. This table presents a comparison of the minimal solution time in minutes (using a decimal notation) obtained by users for solving exercises at each execution level in the two studies. Since the success criteria changed in study B, we also report the minimal time obtained for passing exercises achieving all the goals which is the criteria used in study B. The table also presents the number of users that were able to reach the difficulty levels.

Difficulty Level		Minimal Solution Time (minutes)								
		1	2	3	4	5	6	7	8	9
Study A	80% goals	1.6	2.4	2.68	2.5	4.38	4.04	4.79	3.56	4.3
	100% goals	1.6	2.5	2.68	2.5	4.4	4.25	5.72	3.56	4.32
	N. users	11	11	11	11	11	11	11	11	10
Study B	100% goals	7.0	6.6	7.10	9.74	10.05	14.72	16.4	-	-
	N. users	11	11	11	11	10	6	3	-	-

Table 3. The average exercise generation time (in seconds) that the Trip Generator takes to create 30 exercises at the different difficulty levels including the calls to the planner to verify exercise feasibility.

Average exercise generation time (seconds)									
Difficulty Level	1	2	3	4	5	6	7	8	9
Study B	32	48	48	125	210	633	708	1492	2386

Table 2 presents the minimal solution time to solve an exercise at each level of difficulty and the number of older adults participants that reached a given level. The data in the study A indicates that 10 out of 11 participants reached the highest difficulty levels. In contrast, in study B, only three participants were able to execute the training task at level 7, which is the first level of the difficult stage. This finding demonstrates that the modifications made to increase the difficulty level, such as introducing a spending budget, dinner and lunch goals, path minimisation, and the requirement to execute the plan without errors, made the exercise more challenging. This is also confirmed by the increasing time spent to solve exercises at a given level. In the revised version of Weekend in Rome (Study B), with the exception of the transition between the first and the second difficulty levels, where a learning-related effect can be observed, the progression of difficulty levels is monotonic. Consequently, a notable enhancement has been achieved in comparison to Study A, where a flattening effect of the minimal time required for solving exercises can be observed.

Additionally, Table 3 presents the time required to generate new problems at different levels. This data provides another indicator to measure whether the difficulty of the proposed tasks effectively increases. The Trip Generator always calls the planner to verify that the newly generated exercises are indeed solvable. Thus, if solution plans take longer and are more difficult to find, they are presumably more challenging for users.

Table 4. Means and Standard Deviations for BADS total score and subtests scores, and EPT scores administered at test, retest, and follow-up.

	T1: test (N=11) (N=11)		T2: re-test (N=11)		T3: follow-up (N=10)		ANOVA Time p
	M	(SD)	M	(SD)	M	(SD)	
Battery							
BADS	18.1	(2.4)	19.5	(2.0)	18.4	(2.7)	0.065
BADS: Rule Change	3.5	(0.7)	3.6	(0.5)	3.7	(0.7)	1.000
BADS: Action Plan	3.8	(0.4)	4.0	(0.0)	3.9	(0.3)	0.387
BADS: Search Key	2.2	(1.3)	2.6	(1.0)	2.0	(1.1)	0.327
BADS: Time Estimates	3.0	(0.9)	3.3	(0.8)	2.8	(1.1)	0.150
BADS: Zoo Map	2.3	(0.9)	2.8	(0.9)	2.6	(1.1)	0.460
BADS: Six Elements Test	3.4	(0.9)	3.4	(0.8)	3.4	(0.7)	1.000
EPT	11.7	(1.7)	10.6	(2.2)	11.6	(2.6)	0.180

The main results of the assessment are presented in Table 4, which displays the T1, T2 and T3 BADS's total scores, as well as the BADS subtests scores, and EPT scores. The table shows that the improvement observed at T2 in the re-test phase, was nearly lost at T3 in the follow-up phase.

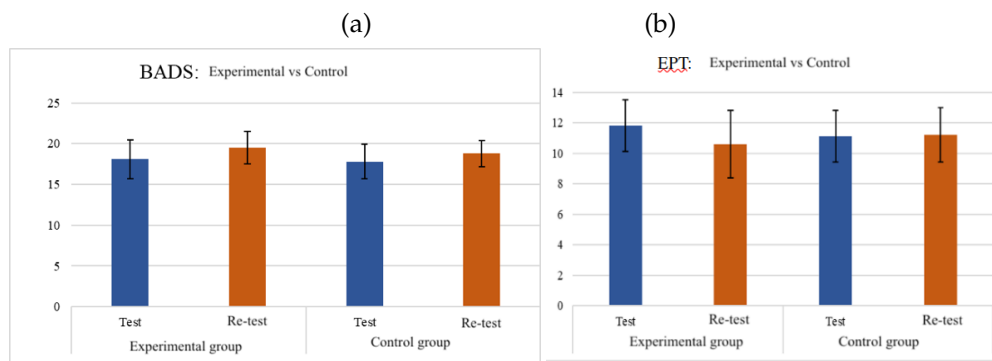


Figure 6. Comparison of the results obtained for the experimental and control groups for the BADS (a) and EPT (b) total scores. The significant improvement for the experimental group on the BADS total score can be observed on the left.

Furthermore, a comparison of the results of the experimental group with those of the control group was carried out (see Figure 6). Separate repeated-measures ANOVAs were conducted on the BADS and EPT scores. The between-subject factor was group (experimental and control group), while the within-subject factor was time (test, retest). No interaction effects reached the significance level. The BADS total score showed a significant simple effect of time [$F(1, 20) = 11.28, p = .003$]. Planned comparisons revealed a significant improvement only for the experimental group ($p = .048$). The effectiveness data are inconclusive. To enhance the reliability of the training results, it would be prudent to expand the size of the experimental group. Additionally, the T1 data indicate a ceiling effect, which implies that the selected tests may have been too straightforward for the participants to demonstrate a change in performance.

4. Discussion

A noteworthy observation from these results is that data on the training exercises at the last two levels could not be obtained. This may be primarily due to the limited duration of the training, which was restricted to eight sessions only. However, the enhanced difficulty level of the proposed exercises in the revised version of Weekend in Rome is also corroborated by the reporting of the time required to generate new problems across different levels, as shown in Table 3. Indeed, the generation of exercises also involves several calls to the planner, which verifies that the newly generated exercises are effectively solvable. Therefore, if solution plans are longer and more difficult to find, it can be presumed that they are more difficult for users. In Study B, participants were aware of the no-error policy for

progressing to the next level, which may have prompted them to allocate enhanced concentration to both the planning and execution of the task.

Given the limited duration of the training and the fact that no one experienced the higher levels of difficulty in the training, it is possible that most participants did not reach their threshold level. However, although we still need to experiment with the training at the highest levels, we have gathered enough information to conclude that the progression of difficulty we have implemented is effective and would support adaptability throughout the training, enabling older adults to tackle problems at the right level of difficulty.

These positive results are also corroborated by the results of the administered usability questionnaire. Indeed, the proportion of respondents who answered affirmatively to the question "Have you encountered any obstacle and/or difficulties during the exercise?" increased from 25% in study A to 45.4% in study B. Similarly, the proportion of respondents who answered affirmatively to the question "The difficulty level of the exercise increased gradually and progressively?" which was assessed with a 5-item Likert scale, increased from 3.4 (an almost neutral score) to 4.3 in Study B. In summary the impression of users was that Version 2.0 of Weekend in Rome presented more challenging tasks with an increasing difficulty level.

In terms of usability and evaluation of ecological features, the results obtained in Study B were similar to those obtained in Study A. This can be seen as a positive result, meaning that moving to a more complex planning system and more complex user interfaces did not have negative effects. However, it also means that further improvements are needed. For example, participants appreciated the introduction of short videos to present historical informations about POIs, but found them repetitive. To address this, we added different videos at different levels of difficulty.

5. Related Work

Pollack [22] identifies three classes of systems that utilise Artificial Intelligence (AI) techniques to support older adults. The first class is that of systems that monitor a person and provide alarms and status reports. The second class of systems is designed to assist older adults in compensating for cognitive impairments. These systems can facilitate the management of daily schedules, the completion of multi-step tasks, the recognition of faces and the localisation of objects. The third class of systems employs AI to provide continuous assessment of the cognitive state of older adults. The systems reported by Pollack represent only a subset of possible applications; in fact, AI can also be used to predict or support older adults who are experiencing cognitive decline [23]. Another noteworthy application is the use of AI techniques to enable older adults to exercise their abilities and to enhance them.

With regard to planning, both experimental and commercial systems, provide tasks to train planning ability. For instance, the implementation of Plan-A-Day presented in [24] or the shopping exercise implement in the Rehacom cognitive training system [25] may be cited as examples. However, the majority of the proposed tasks adopt ad hoc solutions. For instance, the implementation of Plan-A-Day presented in [24] offers only eight fixed problems with increasing difficulty levels, and performance is evaluated based on the solution time rather than the correctness of the plan.

Although less frequent, the use of automatic planning for serious games and training tasks has begun in the last decade, as seen in [10,11,26,27]. However, none of the proposed exercises reaches the complexity of the Weekend in Rome task, which is inherently more difficult than the above examples. It spans multiple days, combines various activities, and utilises an advanced planner.

6. Conclusion

This research effort demonstrates the effectiveness of a participatory design approach in the development of cognitive training tasks for older adults. Following a focus-group study involving older adult to gather requirements for the task [13], two user studies were conducted to refine and tune an initial prototype. The results presented in this paper show that these studies were essential for

improving the features of a cognitive training to train problem-solving abilities. It can be asserted that the utilization of this methodology facilitated the expeditious development of efficacious tasks. enables the aggregation The integration of incremental iterations enabled the aggregation of user insights, culminating in the generation of a better product.

Another equally important contribution concerns the validation of the fine-tuning of the planning task, which the presented results demonstrate to be effective. We have obtained these achievements by exploiting an advanced planner, ENHSP, which adds constraints at the medium stages of difficulty. At this stage, users have to cope with bus schedules that reduce the possible moves and feasible paths. Furthermore, at the difficult stage, minimisation constraints for time and expenses are enforced. New exercises can be created dynamically at a given difficulty level, allowing older adults to train their abilities in a variety of possible scenarios. In conclusion, the results demonstrated that the Weekend in Rome prototype was significantly enhanced, and the effects on the older participants, who engaged in a eight-session training utilising Version 2.0 of the task were encouraging, although inconclusive.

Future work will concern the improvement of the appearance of the training tasks, the addition of features to make it more realistic, the further improvement of the user interface, the introduction of unexpected events, and the addition of support for collaborative sessions. In consideration of the progression of difficulty levels, it was decided to reduce the number of consecutive correct attempts required to advance to the next level. This adjustment would allow more participants to reach the most challenging level, enabling them to train in cost minimisation or path length reduction.

Author Contributions: M.G., S.Z., and F.S. wrote the main manuscript text. M.G. and S.Z. designed and implemented the training task. F.S., D.S. and G.M. designed and conducted the experiments. All authors reviewed the manuscript.

Funding: This work was funded by Velux Foundation with grant.N. 1755 Adaptive and collective intelligent web-based training to enhance problem solving in older people.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committees for Psychological Research of the University of Padova (protocol N. 4371) approved the 17 September 2021.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. United_Nations. Revision of World Population Prospects - Department of Economic Social Affairs Population Division. <https://population.un.org/wpp/>, 2022. Available on line.
2. United_Nations. Revision of World Population Prospects - Department of Economic Social Affairs Population Division. <https://www.un.org/en/global-issues/ageing>, 2019. Available on line.
3. Shishehgar, M.; Kerr, D.; Blake, J. A systematic review of research into how robotic technology can help older people. *Smart Health* **2018**, *7*, 1–18.
4. Beaton, K.; McEvoy, C.; Grimmer, K. Identifying indicators of early functional decline in community-dwelling older people: a review. *Geriatrics & Gerontology International* **2015**, *15*, 133–140.
5. Bezdicek, O.; Červenková, M.; Georgi, H.; Schmand, B.; Hladká, A.; Rulseh, A.; Kopeček, M. Long-term cognitive trajectory and activities of daily living in healthy aging. *The Clinical Neuropsychologist* **2021**, *35*, 1381–1397. doi:10.1080/13854046.2020.1745895.
6. Gross, A.L.; Rebok, G.W.; Unverzagt, F.W.; Willis, S.L.; Brandt, J. Cognitive predictors of everyday functioning in older adults: Results from the ACTIVE cognitive intervention trial. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences* **2011**, *66*, 557–566.
7. Beydoun, M.A.; Beydoun, H.A.; Gamaldo, A.A.; Teel, A.; Zonderman, A.B.; Wang, Y. Epidemiologic studies of modifiable factors associated with cognition and dementia: systematic review and meta-analysis. *BMC Public Health* **2014**, *14*, 1–33.
8. Geda, Y.E.; Silber, T.C.; Roberts, R.O.; Knopman, D.S.; Christianson, T.J.; Pankratz, V. S. and Petersen, R.C. Computer activities, physical exercise, aging, and mild cognitive impairment: a population-based study. *In Mayo Clinic Proceedings* **2012**, *87*, 437–442.

9. Harvey, P.D.; McGurk, S.R.; Mahncke, H.; Wykes, T. Controversies in computerized cognitive training. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **2018**, *3*, 907–915.
10. Baschieri, D.; Gaspari, M.; Zini, F. A planning-based serious game for cognitive rehabilitation in multiple sclerosis. Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good; ACM: New York, NY, USA, 2018; p. 214–219. doi:10.1145/3284869.3284916.
11. Gaspari, M.; Pinardi, F.; Signorello, D.; Stablum, F.; Zuppiroli, S. Automatic planning in cognitive training: application to multiple sclerosis. *Human-Computer interaction* **2003**, *38*, 173–196.
12. Gaspari, M.; Donnici, M. Weekend in Rome: a cognitive training exercise based on planning. SAT@ SMC, 2019, pp. 37–41.
13. Cipolletta, S.; Signorello, D.; Zuppiroli, S.; A., H.; Ballhausen, N. Mioni, G.; Kliegel, M.; Gaspari, M.; Stablum, F. A focus group study for the design of a web-based tool for improving problem-solving in older adults. *European Journal of Ageing - in press* **2024**, *21*, 1–12. doi:10.1007/s10433-024-00814-0.
14. Lu, M.H.; Lin, W.; Yueh, H.P. Development and evaluation of a cognitive training game for older people: a design-based approach. *Frontiers in Psychology* **2017**, *8*, 1837.
15. Haslum, P.; Lipovetzky, N.; Magazzeni, D.; Muise, C.; Brachman, R.; Rossi, F.; Stone, P. *An introduction to the planning domain definition language*; Vol. 13, Morgan & Claypool Publishers: San Rafael, California, 2019.
16. Pellier, D.; Fiorino, H. PDDL4J: a planning domain description library for java. *Journal of Experimental & Theoretical Artificial Intelligence* **2018**, *30*, 143–176. doi:10.1080/0952813X.2017.1409278.
17. Scala, E.; Haslum, P.; Thiébaux, S.; Ramirez, M. Interval-based relaxation for general numeric planning. ECAI 2016; IOS Press: Amsterdam, NL, 2016; pp. 655–663.
18. Wilson, B.A.; Alderman, N.; Burgess, P.W.; Emslie, H.; Evans, J. *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*; Thames Valley Test Company: Bury St. Edmunds, Suffolk UK, 1996.
19. Antonucci, G.; Spitoni, G.; Orsini, A.; D'Olimpio, F.; Cantagallo, A. *A. Behavioural Assessment of the Dysexecutive Syndrome. La batteria per la valutazione dei deficit delle funzioni esecutive. [Italian adaptation]*; Giunti O.S.: Firenze, Italy, 2014.
20. Willis, S.L.; Marsiske, M. *Manual for the everyday problems test*; University Park: Department of Human Development and Family Studies, Pennsylvania State University: Pennsylvania, USA, 1993.
21. Borella, E.; Cantarella, A.; Carbone, E.; Zavagnin, M.; De Beni, R. *Quotidiana-mente: La valutazione dell'autonomia funzionale e dell'auto-percezione di fallimenti cognitivi in adulti-anziani*; FrancoAngeli: Milano, Italy, 2017.
22. Pollack, M.E. Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment. *AI magazine* **2005**, *26*, 9–9.
23. Graham, S.A.; Lee, E.E.; Jeste, D.V.; Van Patten, R.; Twamley, E.W.; Nebeker, C.; Yamada, Y.; Kim, H.C.; Depp, C.A. Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry research* **2020**, *284*, 112732.
24. Holt, D.V.; Rodewald, K.; Rentrop, M.; Funke, J.; Weisbrod, M.; Kaiser, S. The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of the International Neuropsychological Society* **2011**, *17*, 327–335. doi:10.1017/S1355617710001712.
25. López-Martínez, Á.; Santiago-Ramajo, S.; Caracuel, A.; Valls-Serrano, C.; Hornos, M.J.; Rodríguez-Fórtiz, M.J. Game of gifts purchase: Computer-based training of executive functions for the elderly. 2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH). IEEE, 2011, pp. 1–8.
26. Menif, A.; Guettier, C.; & Cazenave, T. Planning and execution control architecture for infantry serious gaming. Proceeding of the 3rd International Planning in Games Workshop, ICAPS 2013; Borrajo, D.; Kambhampati, S.; Oddi, A.; Fratini, S., Eds.; AAAI Press: Washington, DC USA, 2013.
27. Do, M.; Tran, M. PBlocksworld: An iPad puzzle game. Proceeding of the 3rd International Planning in Games Workshop, ICAPS 2013; Borrajo, D.; Kambhampati, S.; Oddi, A.; Fratini, S., Eds.; AAAI Press: Washington, DC USA, 2013; pp. 35–39.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.