MASTER

Impact of Interruptions on Older Adults' Activities of Daily Living
Effects on Task Duration, Errors and Strategy

Giesen, Rinke J.

Award date:
2024

PO Box 513
5600 MB Eindhoven
The Netherlands
tue.nl

DEPARTMENT OF INDUSTRIAL ENGINEERING & INNOVATION SCIENCES

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

MASTER THESIS

Impact of Interruptions on Older Adults' Activities of Daily Living: Effects on Task Duration, Errors and Strategy

Author: R.J. Giesen
Student Number: 1328298
Date: 25-01-2024

DEPARTMENT OF INDUSTRIAL ENGINEERING & INNOVATION SCIENCES

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Eindhoven, 25-01-2024

Impact of Interruptions on Older Adults' Activities of Daily Living: Effects on Task Duration, Errors and Strategy

by R.J. Giesen

1328298

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

First supervisor:
Dr. Linghan Zhang

Second supervisor:
Prof. Dr. Wijnand IJsselsteijn

Third Assessor
Dr. Ir. Raymond Cuijpers

Abstract

This study was designed to gain insights into how older adults react to an interruption during the performance of an activity of daily living (ADL). Participants were required to make a packed lunch (Schwartz et al., 2002) in the Usability lab which contained a naturalistic living room setting. The interruption occurred around one-third of the task and was intended to disengage participants from the execution of the ADL. The main variables of interest were time on task, type and frequency of errors made, and subjective workload. Moreover, participants were interviewed to get an understanding of their strategy to deal with the interruption. 20 healthy older adults and one person with dementia participated in the study, all were aged 65 years or older. Results showed that although the interruption succeeded in mentally disconnecting participants from the NAT, it did not disrupt their planned and actual performance. Moreover, participants indicated that they did not use techniques to remember the instructions of the NAT. Time on task and number of errors could not be predicted by participants' subjective workload, cognitive functioning, and daily life independence on IADLs. This, together with the absence of a speed-accuracy tradeoff, is likely due to the high familiarity of the performed ADL task. The participant with dementia had the longest time on task, made the highest number of errors, and had the second-highest subjective workload. This study showed that familiarity with a task can affect performance, especially when the current task requires different requirements than usual. This knowledge can be used to improve the design of assistive technologies as it stresses the importance of learning the habitual ADL execution of people to help them in cases when the habitual response is not correct. The lack of participants with dementia limited the study in investigating differences between healthy older adults and people with dementia. Future studies could look at the possibility of using AI to measure ADL performance at home, instead of in the lab to increase the ecological validity of the study.

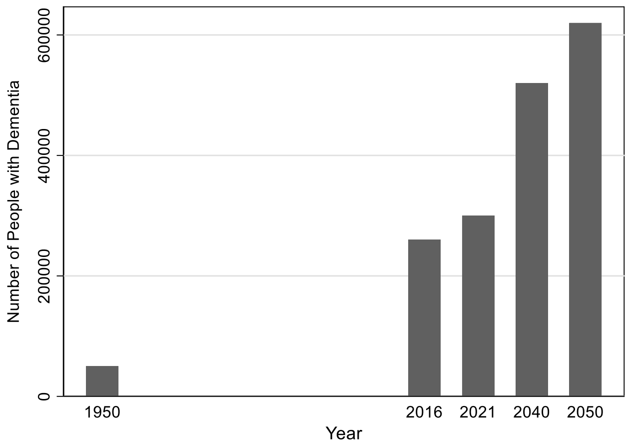
Introduction

With the implementation of the Long Term Care Act (in Dutch “wet langdurige zorg”) in 2015, the Dutch healthcare policy regarding older adults has changed drastically (Kromhout et al., 2018; Verkooijen, 2020). One of the main goals of this Act was to keep future healthcare costs affordable by letting older adults live independently at home for as long as possible (Verkooijen, 2020). That is why, from that moment on, to qualify for a nursing home, a person must require intensive care (Ministry of Health Welfare and Sport, n.d.; Verbeek-Oudijk & Campen, 2017). This Act was intended to amplify the already ongoing trend of more older adults living independently at home (Verbeek-Oudijk & Campen, 2017). This appeared to work as the number of older adults living in care homes decreased from 115.000 in 2015 to 113.000 in 2019 (Verbeek-Oudijk & Koper, 2021). However, at the same time, the number of people living at home with disabilities, like dementia, was rising.

In 2018, it was estimated that 205.000 people in the Netherlands lived at home while suffering from dementia (Homan et al., 2018). In 2021, this number had already increased by almost 16% to 237.000 (Alzheimer Nederland, 2021). In the same year, it was estimated that in total 300.000 people in the Netherlands suffered from dementia (Alzheimer Nederland, 2021). It is expected that this number will rise substantially in the future (Figure 1; Alzheimer Nederland, 2021; Patterson, 2018), as the Dutch population is aging, the number of people that are 65 years and older will rise from around 3.6 million in 2023 to 4.8 million in 2050 (Centraal Bureau voor de Statistiek, n.d., 2020) and because the prevalence of dementia, and many other chronic diseases, increases with age (Berr et al., 2005; Denton & Spencer, 2010). In Western Europe, dementia prevalence numbers go from 1.6% for people between 60-64 to 43.1% for people older than 90 years (Prince et al., 2013).

Figure 1

The Number of People with Dementia in the Netherlands over the Years



Note. Numbers taken from Alzheimer Netherlands (2021)

To live independently at home, proper execution of Activities of Daily Living (ADL) is crucial (Edemekong et al., 2019). One group that has difficulties with executing ADLs is people with dementia (Beaver et al., 2019; Waite et al., 2000). There are various types of dementia and although they share some features, like problems with memory (Thompson et al., 2010), they also have some more specific, but not exclusive, symptoms as people with Lewy Body dementia were found to have more hallucinations, whereas people with vascular dementia were found to be more paranoid and delusional (Chiu et al., 2006). Additionally, vascular dementia, Alzheimer’s and particularly Lewy Body dementia were found to relate to abnormal gait, characterized by ataxia and slow movements (Waite et al., 2000) Next to this, Alzheimer’s disease and frontotemporal dementia were associated with a decline in cognitive functioning (Nedjam et al., 2004). Frontotemporal dementia was also related to more socially inappropriate behavior (Neary et al., 1998). (Pre-clinical) Alzheimer’s, on the other hand, was found to relate to different aspects of memory (Jones et al., 2006; Kopelman, 1985). Performing ADLs requires executive functioning and memory which are mental processes that are often impaired by people with dementia creating difficulties with ADL execution (Beaver et al., 2019; Cahn-Weiner et al., 2002; Waite et al., 2000; Woods et al., 2012).

Assistive technologies have been developed to facilitate people with dementia who need help to complete essential ADLs (Astell et al., 2019; Mihailidis et al., 2008). However, there are still large knowledge gaps present that prevent the proper design of assistive technologies. One of the research gaps is regarding the effect of interruptions on the execution of ADLs. That is why this study had the goal of investigating how older adults react to an interruption while making a packed lunch, a task sequence that combines a number of ADLs. The outcome of this study was transformed into recommendations for assistive technologies.

Dementia and its Influence on ADL Performance

ADLs can be divided into two categories: basic ADLs (BADLs) and instrumental ADLs (IADLs; Edemekong et al., 2019). BADLs like managing personal hygiene, eating, and dressing require basic physical skills and are affected by motor impairments (Boyle et al., 2002; Edemekong et al., 2019). IADLs, on the other hand, require a higher cognitive performance as these skills are about managing finances, meal preparation, house cleaning, and medication (Edemekong et al., 2019). IADLs are associated with cognitive abilities, for example, baseline memory and executive functioning (Cahn-Weiner et al., 2007). As dementia is characterized by problems with executive functioning and memory (Jones et al., 2006; Kopelman, 1985; Nedjam et al., 2004; Thompson et al., 2010), it is not surprising that these people have more difficulties executing everyday tasks and have in general a worse performance than healthy older adults (HOA; Beaver et al., 2019).

The two categories of ADL are not only impacted by different types of impairments but they were also linked to different stages of dementia. For instance, BADLs are more related to motor impairments (Boyle et al., 2002) and are impaired in the more severe stages of dementia (Shimokihara et al., 2022). IADLs, on the other hand, are linked to the early stages of dementia. For example, the longitudinal study from Pérès et al. (2008) showed that future dementia patients had already worse IADL scores ten years before the official clinical diagnosis compared to people who did not develop the disease. This poses the question of whether IADL assessment could be used for the diagnosis of early-stage dementia (De Lepeleire et al., 2004). The question is then how to assess ADL performance in naturalistic settings.

Measurements of ADL Performance

Questionnaires

There are mainly two types of ADL performance measurements. The first is questionnaires developed to understand how well a person can execute ADLs. These questionnaires can be further classified as self- and informant reports based on the information sources. For example, Lawton and Brody (1969) proposed the Instrumental ADL scale widely used to measure IADL performance; see for example Edemekong et al. (2019) or Villeponteaux et al. (1998). The IADL scale uses 8 items, like household, managing finances and using transportation, to rate a person's performance on IADLs (Lawton & Brody, 1969). Another example is the Barthel Index (BI), which has been widely used to measure BADL performance; see for example Lai et al. (2002) or Saito et al. (2017). Ten items, like bathing, dressing, and eating, are used in the BI to score a person's BADL performance (Hartigan, 2007).

Both the BI and IADL questionnaires ask either the participant or a caregiver to elaborate on the capabilities of conducting certain ADLs independently. Subsequently, the ADL performance per category will be rated, by scoring either zero or one point in which one point indicates that the participant can still do this ADL independently (Bouwstra et al., 2019; Lawton & Brody, 1969). Whereas in the IADL Questionnaire, the only scoring options are either one or zero points, the BI also has some categories in which 2 or 3 points can be given. In the latter case, the higher the number of points, the more independently the participant can execute the ADL (Bouwstra et al., 2019). In comparison to the previously two discussed questionnaires, the Amsterdam IADL Questionnaire only asks the caregivers to evaluate the IADL performance of the participant (Sikkes et al., 2012). It includes 70 items such as “Did he/she carry out household duties in the past 4 weeks?” (Sikkes et al., 2012; Sikkes et al., 2017, p.1). The main goal of the Amsterdam IADL Questionnaire is to detect early(-onset) dementia (Sikkes et al., 2012).

Questionnaires are simple to administer and provide reasonable accuracy of real-world ADL performance as they offer the possibility of multiple observations (Schmitter-Edgecombe et al., 2011). On the other hand, using questionnaires has disadvantages as well as they can be biased by the current mood or affective state of the participant. Research showed that when inducing participants with happy, sad, or neutral affective states or moods, answers on questionnaires differed between groups (Gillham et al., 2007; Pinilla et al., 2020). Moreover, the environmental context also influences participants. Research showed that different desk sizes influenced the answers on a questionnaire measuring perceived comfort and intended self-disclosure of patients in a hospital setting (Okken et al., 2012).

Another limitation of questionnaires is that they cannot be used to analyze the exact process and the difficulties that arise when performing ADLs (Giovannetti et al., 2019). Lastly, research suggested that people with dementia have less awareness of their limitations, which can cause discrepancies between their self and other-reported capabilities (Farias et al., 2005). A solution for the subjectiveness of questionnaires would be to let people execute

ADLs in a controlled environment, like a lab, and monitor them; this class of measures is called performance-based measures and is the second way to measure ADL performance.

Performance-based measures

Performance-based measures require participants to solve realistic problems, like calculating the cost of transportation on paper (EPT; Schmitter-Edgecombe et al., 2011), looking up a specific phone number (OTDL-R; Diehl et al., 2005), or conducting daily tasks, such as making coffee and lunch packs with real-life materials (NAT; Schwartz et al. 2002).

In this study, the method is based on the Naturalistic Action Test (NAT), which can be used to measure the performance of a person working on everyday activities (Schwartz et al., 2002). The NAT is composed of 3 tasks and 20 steps in total, which can be time- and effort-consuming and less feasible for people with dementia. That is why some studies, like Divers et al. (2021) and Giovannetti et al. (2010) only include parts of the NAT, which can be justified by the fact that Schwartz et al. (2002) found that 88.5% of the participants in their validation study who scored low on the first task also scored low on the second and third tasks. At this moment, the NAT is often used in lab-based settings, see for examples the studies from Rodríguez-Bailón et al. (2017), or Giovannetti et al. (2008) However, there are also promising efforts made to create a VR version of the NAT (Giovannetti et al., 2019).

To quantify and compare performance across participants, researchers defined several errors that participants can make during performance-based measurements. Most important are omission errors which occur when omitting a crucial step for successfully ending the task (Giovannetti et al., 2010; Rodríguez-Bailón et al., 2017; Weakley & Schmitter-Edgecombe, 2019). Next to this, errors regarding the sequence of steps, the substitution of the target object for another object and the addition of steps that are not needed for the successful ending of the target task are mentioned regularly (Giovannetti et al., 2010; Rodríguez-Bailón et al., 2017; Weakley & Schmitter-Edgecombe, 2019). Some studies also registered tool omissions, which happen when the appropriate tool necessary for task completion is not used (Giovannetti et al., 2010; Rodríguez-Bailón et al., 2017).

Additionally, to the errors described above, this study also considered micro errors. Micro errors entail reaching toward, touching, or moving objects that are not necessary for the current task and can be divided into three levels indicating the severity of the error made (Divers et al., 2021). The level of micro error is indicative of the performance of the error monitoring system of participants, which describes how fast participants detect and correct their errors (Divers et al., 2021). That is why reaching for a wrong object is only a mild error, as the error monitoring system is efficient enough to detect and correct the error on time. Touching a wrong object is a moderate micro error and the actual movement of a wrong object is a severe micro error and is indicative of a less efficient error monitoring system (Divers et al., 2021).

Compared with questionnaires, performance-based measures are observable, standardized and objective (Giovannetti et al., 2019), making them unbiased. That is why performance-based measurement solves many of the limitations created by the use of questionnaires. However, since they are primarily administered in controlled lab environments, some argue that they are not equivalent to direct observations of ADLs in naturalistic settings like the participants' homes (see for example Rodríguez-Bailón et al. (2017). This is especially relevant, as lab experiments typically leave out factors that might otherwise affect ADL performance, such as interruptions.

Interruptions and their relation to ADL performance

Many researchers have explored how differences in the timing of interruptions affect individuals' task performances. Research showed that resumption lags, which is defined as the time a person needs to (re)start on the task after an interruption, and completion times were longer when interruptions appeared mid-task compared to their appearance before a task (Altmann & Trafton, 2004; Bailey & Konstan, 2006; Monk et al., 2004). At the same time, it was also found that for older adults this effect was overshadowed by age-related effects, as older adults had a longer resumption lag overall (Monk et al., 2004). This might indicate that interruptions, in general, are very disruptive for older adults, irrespective of the timing.

The timing of interruptions also influences the reaction of the receiver as research showed that people deal with them faster when they appear during a task (Bailey & Konstan, 2006). Moreover, research showed that mood is affected by the moment of interruption appearance as participants experienced less positive moods and more annoyance and frustration when an interruption occurred within a task (Adamczyk & Bailey, 2004; Bailey & Konstan, 2006; Zijlstra et al., 1999). Therefore, it can be concluded that interruptions and the moment of their appearance do affect people and their execution of sequential tasks.

In daily life, interruptions, like receiving calls or messages, happen on a daily basis and should be taken into account when executing ADLs. However, there is to our knowledge only one paper that investigated how mild cognitively impaired older adults were impacted by an unexpected interruption while doing an ADL. This paper, from Weakley and Schmitter-Edgecombe (2019) used people with mild cognitive impairments (MCI) as participants. Literature defines this group as people who have mild cognitive impairments that are worse than the ones resulting from normal aging but are not as bad as the cognitive impairments resulting from dementia (Petersen, 2004). A literature review from Jekel et al. (2015) showed that performance on IADLs of people with MCI is generally in between HOA and people with dementia.

Weakley & Schmitter-Edgecombe (2019) designed an experiment involving a sequential task structure involving two sets of activities, either the making of oatmeal or lemonade. Both task sequences consisted of 20 steps that people executed in an apartment located on campus. Their experiment consisted of two conditions of which one involved a short interruption that happened immediately after the first step. The interruption started by the researchers calling the participant with the message they forgot their stopwatch and the question whether the participant could retrieve it and bring it to the foyer at the bottom of the stairs. It was chosen to start the interruption after the participant executed the first step of the sequence to retain maximum control over the total number of steps left and the timestamp of the interruption. However, this raises the question of what the effect is of an interruption later in the sequence. This is a question the current research tried to answer.

The experimental results from Weakley and Schmitter-Edgecombe (2019) showed that people with MCI took longer to complete the task and made more substitution errors but the same amount of omission errors as people with MCI in the *no interruption* condition. HOA, on the other hand, had the same completion time but did make more omission errors compared to the HOA in the other condition. The authors believed that this might be because the two groups of participants adopted different speed-accuracy trade-off strategies. They suggested the HOA focused on speed whereas the people with MCI focused on accuracy. However, as they did not include a self-report component in their studies, they could not verify whether participants consciously altered their strategy after the interruption. The current research added such components to investigate to what extent conscious strategies were present in sequential tasks and whether they changed after an interruption.

Interruptions can also be used for positive purposes. For example, a study by Caroux et al. (2020) used a different viewpoint and investigated how interrupting notifications impacted older adults when working on an ADL. The goal of the study was to get more insight into how assistive technology should be designed. It was found that critical notifications were much more accepted than non-critical ones. Moreover, the intrusiveness of the notification tone should match the urgency of the content for better acceptance (Caroux et al., 2020). Also interesting was the finding that notifications were in general less accepted when participants worked on IADLs that mainly required cognitive resources, like bookkeeping (Caroux et al., 2020). Therefore, it would be concluded that assistive technology should be context-aware to present the interruption at the best possible time.

The study from Caroux et al. (2020) fits in line with the argument from Astell et al. (2019) that technology can be used for caregiving purposes for people with dementia. Technology can intervene with prompts and reminders to support people with dementia who live at home and need help, socially, physically, or cognitively, to complete essential ADLs (Astell et al., 2019). However, if not designed properly, assistive technology can also, paradoxically, become an interruption in itself. At this moment, knowledge is lacking as it is largely unknown how interruptions impact people with dementia or MCI on their execution of ADLs. Research is necessary for the following two reasons. First, it is important to understand what kind of errors people with dementia or MCI make and why, after being interrupted. This knowledge could be used in the design of assistive technologies to help people recover from interruptions by preventing them from making these errors. Secondly, it is valuable to know how the timing of interruptions impacts ADL performance. This way, a more appropriate timing of the prompts of assistive technologies could be designed as well. As scientific research is lacking, not all questions can be answered in one study. The current study focused on the impact of an interruption and transformed the findings into recommendations regarding the design of assistive technology.

This study

In a time in which older adults are encouraged to live at home for as long as possible, sometimes while having dementia, assistive technologies may add significant value in helping people stay independent for longer. Unfortunately, with little research available on the role of interruptions in ALDs, still a lot is unknown about how older adults, including people with MCI or dementia, react to interruptions during ADLs. As interruptions are a part and parcel of everyday life, and likely to add to the cognitive load of ADL, this lack of knowledge can prevent designers from making an appropriate technological tool to help this group. Therefore, this study was organized to gain more insight into how older adults react to interruptions. The research outcomes were used to formulate useful recommendations for assistive technologies.

At this moment, research is still inconclusive about the effect of an interruption happening in a later stage of an ADL task sequence. Also for older adults, research about the timing of interruptions showed mixed results. The study by Monk et al. (2004) even suggested that timing does not matter for older adults, because of age-related effects. The current research tried to fill this knowledge gap by exploring the effect of an interruption that happened later in the task sequence.

As the interruption in this study was later in the task sequence, there was more time between the instruction and the interruption compared to the study from Weakley and Schmitter-Edgecombe (2019). That is why it was expected that the results found in Weakley and Schmitter-Edgecombe (2019) would be amplified in this study. Unlike, the study from Weakley and Schmitter-Edgecombe (2019), the current study did not split participants into groups, based on their cognitive functioning. Instead, the current study used the measured cognitive functioning scores as a predictor in regression analyses. Based on the study from

Weakley and Schmitter-Edgecombe (2019), it was expected that the lower people's cognitive functioning, the more they experienced task interference by the interruption. That is why it was hypothesized that after the interruption people with lower cognitive functioning have longer completion times, longer resumption lags and more (substitution) errors compared to people with higher cognitive functioning.

Additionally, earlier research did not include subjective data that might explain whether participants consciously used different strategies to deal with the interruptions. That is why this research included both a subjective workload measurement and a short interview to gain insight into how subjective workload relates to performance and how participants evaluated their own strategies, before and after the interruption. It was hypothesized that people who indicated a higher subjective workload would take longer to complete the task and make more errors in general. Moreover, it was expected that participants who had consciously a strategy in mind had in general a higher performance on the ADL task than participants who did not.

Finally, the IADL scale from Lawton and Brody (1969) was administered to see how performance at the NAT would relate to the IADL score. It was hypothesized that people who have a lower IADL score also made more errors on the NAT as they are less independent in doing ADLs.

Method

Design

The study took the form of a post-hoc experimental design. The data for this study were collected by observing participants performing an ADL task based on the NAT (Schwartz et al., 2002) and by conducting questionnaires and interviews. During the NAT, all participants experienced the same interruption. The main variables of interest were the completion time of the NAT and the frequency and type of errors participants made. These variables of interest were analyzed using multilevel regression with the Mini-Mental State Exam (MMSE) score as the main predictor. The MMSE is an 11-item questionnaire that measures cognitive functioning. Data collection took place in November and December 2023 in the Usability lab at the Eindhoven University of Technology and in the residential care center Savant van Lenthof in Nuenen.

Participants

In total, 20 older adults without an official diagnosis of dementia participated in the study, of which 11 were female. Age ranged from 65-89 (*M*= 73.85, *SD* = 6.05). The mean score on the MMSE was 28.45 (26-30, *SD*=1.05). More than half of the participants (11) had a University of Applied Science degree. Four participants did not follow an education other than high school. Three participants had a practical education, and only one participant had a University degree. The educational background of one participant was unknown. More than half of the participants had two children (11 participants), although having three children was also frequently found (5 participants). Participants also indicated to have zero (1 participant), one (2 participants), or four (1 participant) children. Furthermore, only two participants indicated to be caretakers. These questions were asked because it was expected that when a participant had kids or was a caregiver, they were more familiar with packing a lunch. Five participants indicated that they suspected that the interruption was on purpose. It is possible that more participants thought so, but did not specify this directly. No participant worked in the catering industry or any industry relating to the ADL task of making a packed lunch.

Additionally, to the 20 participants discussed above, four older adults with an official diagnosis of dementia started the study. Unfortunately, only one of these participants successfully finished it. One of the participants did not start the NAT as the explanation was overwhelming her and caused her discomfort. Another participant had the same overwhelming experience after starting the NAT. For both participants, the study was stopped, and participants were comforted. A third participant did not fully understand the NAT, resulting in unusable data. Moreover, the attention span of this participant was too short to complete the full study, which is why the study was stopped after the interview. This resulted in only one participant with dementia who completed the study. This 90-year-old male did not remember the exact form of dementia he suffered from, or what his educational background was. He has one child and scored 25 points on the MMSE.

All participants without an official diagnosis of dementia were 65 years or older and lived independently at home. The reason why only participants aged 65 or older were collected was to include participants who already experienced some mild age-related cognitive decline to yield some sensitivity to the interruption. Moreover, research mostly uses this age range (Berr et al., 2005) and additionally, the age of 65 is used as a cut-off point for early-onset dementia as well (Lambert et al., 2014; Patterson, 2018).

Recruitment happened mostly via the ARCHIE database (16 participants). Yet, it was noticed that most of these participants were healthy older adults. That is why, to create a more diverse group of participants, also a lot of effort was spent on collecting participants via other ways, for example, by using personal networks. Next to this, a flyer (Appendix A.1) promoting the experiment was made for recruitment purposes and this was shared via LinkedIn. Furthermore, four Alzheimer Café were attended, two in Eindhoven (Alzheimer Nederland, n.d.-b) and two in Oirschot (Alzheimer Nederland, n.d.-a). On all occasions, the flyer was handed out and the experiment was explained to the attendees. Moreover, the flyer was shared in the newsletter of the Alzheimer café in Oirschot, reaching 311 possible participants. Next to this, the flyer was shared with all the nurse case managers in the regions of Eindhoven and Oirschot.

The goal was to have the most diverse participant group possible. That is why the main inclusion criteria were that participants needed to be 65 years or older, living independently at home, potentially with a partner. Moreover, they needed to be physically and mentally able to fulfill the task, meaning that they needed to be able to pour in a drink, make a sandwich and slice a piece of gingerbread. They also needed to be able to converse and answer the questions of the questionnaires, although the experimenter could help them with reading the questionnaires. Lastly, they needed to be able to travel to the university independently to participate in the study.

It appeared to be very hard to recruit participants with official diagnoses of dementia who are living at home, especially in this short time period. Besides, this group of people often had difficulties coming to the university. That is why this inclusion criteria was dropped, in the hope of enlarging the possibility of recruiting participants with dementia. Via contacts at the Expertise Center Dementia & Technology, it was possible to conduct the study at the residential care center Savant van Lenthof in Nuenen.

Before the beginning of the study, participants read and signed the informed consent forms and they had the opportunity to ask questions. Participants were informed that they could stop the experiment at any time possible and that they could withdraw their permission for the use of their data at a later moment as well. A monetary reward was given to the participants after completing the experiment.

Materials and settings:

The experiment was conducted in the Usability lab in the Atlas building at the Eindhoven University of Technology. Figure 2 shows this lab, which is furnished as a living room which was thought to create the most naturalistic setting for the experiment.

Figure 2

The Usability Lab used for the Study

A Sony HDR-CX240 video camera was used to record the NAT. This camera was placed on a tripod on the side of the participant and was directed at the table with the objects necessary for the NAT. The left picture in Figure 3 shows an example of the recording area of this camera. The tripod was used to create an angle from above the participant which ensured a good view of their hand movement, while at the same time, avoiding the recording of identifiable information about their facial areas. Next to this, the video system of the usability lab was used as well to video record the NAT. This camera was placed on the ceiling behind the participant and was also directed at the table. The right picture of Figure 3 shows an example of the recording area of this camera.

Figure 3

Example of the Recording Areas of the two used Cameras

The experiments conducted at the residential care center Savant van Lenthof took place in a “neighborhood” room which is a type of living room in which residents can meet each other (Figure 4). This room was chosen as it created a welcoming, relaxed atmosphere similar to the Usability lab. The tripod with the Sony camera was again used to record the NAT and Figure 5 shows the recording area of this camera.

Figure 4

The Neighborhood Room in Savant van Lenthof

Figure 5

Example of the Recording Area of the Camera Used in Savant van Lenthof

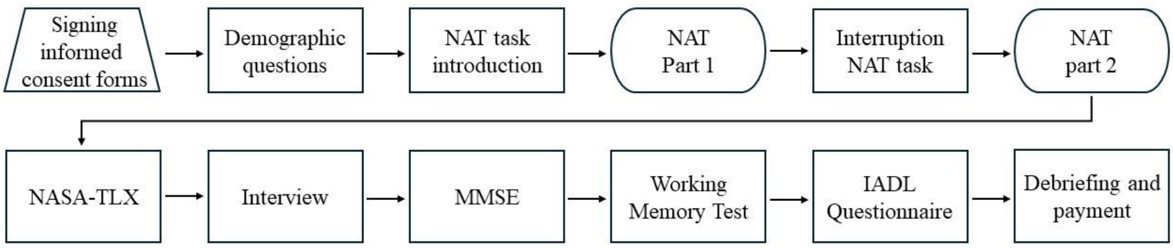
The Zoom H2next handy recorder was used to record the interview. This was done by placing the recorder on the table between the experimenter and the participant during the interview.

Procedure

The study follows the workflow as shown in Figure 6. Before entering the lab, participants would come into an entrance room, where they registered for the study and signed their informed consent. By signing the form, participants agreed with their data being measured and used. Moreover, they agreed with the video recording of the NAT and the audio recording of the interview. After signing the informed consent, the caregiver of the participant, if present, was brought to a control room where a TV screen with live video footage and audio from the lab was present. This way, caregivers could follow everything that happened without interfering with the experiment.

Figure 6

Flowchart of the Experimental Procedure



Note. Entrance room indicated by; . Table 1 indicated by; .
Table 2 (used only for the NAT task) indicated by; .

Following the workflow from Figure 6, participants were brought into the lab after they signed their informed consent form. In the lab, they were placed at the first table, which is the right table in Figure 2. At this table, the procedure of the study was again explained to them and, because it concerned a vulnerable participant group, it was again stressed that participants could always ask questions and stop their participation. Thereafter, demographic questions (e.g. gender, age) were asked (Figure 6).

After answering these questions, the cameras were turned on and the NAT was explained to the participant (Figure 6). The participants were told to make a packed lunch at a second table, which is the left table in Figure 2. They were told that on this table all the necessary objects would stand. It was instructed that participants needed to imagine being at home while making the packed lunch. To make the packed lunch, they were instructed to start by filling a steel thermos bottle with orange juice and then close the bottle. Next, they were told to grab one slice of bread, butter it and spread peanut butter on it. After this, they needed to slice the bread into two halves, place the two halves on top of each other and place it in the lunch box. Then they should again grab a slice of bread, butter it and put cheese on it. Again, they should slice the bread into two halves, place those halves on top of each other and place the sandwich into the lunch box. Finally, participants were told to slice a piece of gingerbread and put this in the lunch box as well. Then the lunch box should be closed and both the lunch box and the steel thermos bottle should be packed in the backpack to end the task.

All 44 steps that were necessary for successful NAT completion can be found in Table 1. The NAT consisted of three different tasks. The columns of Table 1 show these different tasks and the specific steps that are necessary to complete them. The first task (drink) consists of ten steps, the second (sandwich) of 21 steps, and the last task (snack) consists of seven steps. Packing the backpack with the lunch box and the thermos bottle consists of six steps, but was not regarded as a task. To make the instructions clearer, schematic drawings of the three tasks were shown to the participants while explaining the instructions. These three drawings can be seen in Appendix B.1.

After hearing the instructions, the participant was asked to repeat it. When this was unsuccessful, the experimenter explained the task once more and subsequently asked the participant again to repeat the instructions. This procedure was done until the participant could fully reproduce the full instructions. In all cases, after the participant successfully repeated all the instructions, the experimenter repeated the instructions once more, making sure that every participant heard the instructions at least twice. It was then also mentioned that all objects that were necessary for task completion would be present on the table and it was asked to only use the items necessary for task completion. When there were no questions from the side of the participant, he or she was guided to a second table on which all the objects were placed. Following the workflow (Figure 6), participants were notified that when they arrived at the table they were allowed to start.

Table 1

All the Steps Necessary for Successful Task Completion

Steps	Task			
	Drink	Sandwich	Snack	Backpack
1	Grab the orange juice	Grab the bread	Grab the gingerbread	Grab the backpack
2	Open the orange juice	Open the bread	Open the gingerbread	Grab the lunch box
3	Grab the steel thermos bottle	Grab one slice of bread and put it on the plate	Grab a knife	Put the lunch box in the backpack
4	Pour the orange juice	Grab the butter	Slice one piece of gingerbread	Grab the thermos bottle
5	Find the lid of the thermos bottle	Open the butter	Put the gingerbread in the lunch box	Put the thermos bottle in the backpack
6	Place the lid on the thermos bottle	Grab a knife	Grab the lid of the lunch box	Close the backpack
7	Tighten the lid on the thermos bottle	Butter the slice	Close the lunch box	
8	Find the cap of the thermos bottle	Grab the peanut butter		
9	Place the cab on the thermos bottle	Open the peanut butter		

Steps	Task			
	Drink	Sandwich	Snack	Backpack
10	Tighten the cab on the thermos bottle	Put the peanut butter on the slice		
11		Slice the slice of bread into two halves		
12		Place the two halves on top of each other		
13		Put the sandwich in the lunch box		
14		Grab one slice of bread and put it on the plate		
15		Butter the slice		
16		Grab the cheese		
17		Open the cheese		
18		Put the cheese on the slice		
19		Slice the slice of bread into two halves		
20		Place the two halves on top of each other		
21		Put the sandwich in the lunch box		

The interruption (Figure 6) was located after the 11th step of the task sequence, which corresponds to the moment after the first step of making the sandwiches, for example grabbing the first slice of bread. The step order could differ between participants as they are free to follow their own order of steps within the three tasks. The interruption started with the experimenter informing the participants that some questions were forgotten to be asked before the NAT started and that these should be asked before continuing with the NAT. The researcher stressed that these questions had to be answered right away and guided the participants back to the first table to do so. After all the questions were answered, the participants were thanked and informed that they could go back and continue with NAT, which is also indicated in Figure 6.

Participants might run into difficulties when executing the ADL task. That is why nondirective cues were given when participants stopped working or making progress on the tasks. An example of this cue is, "*You can continue your task*" (Giovannetti et al., 2010), which was given around 30 seconds after participants were visually stuck. After 30 more seconds, participants were reminded that everything they needed for task completion was on the table. When 30 more seconds passed and participants were still stuck, they were reminded of their task without details. For example, they were reminded that they needed to make two sandwiches. When in total, two minutes had elapsed in which the participants were stuck, the full instruction of one task was given. For example, "*Your task is to make two sandwiches, one with butter and peanut butter and one with butter and cheese. You need to store them in a lunchbox*". It was chosen to not stop the NAT as it was thought to be a waste of data as participant collection was expected to be hard. It was noted down when participants received which cue, and this was taken into account for the data analysis.

After the NAT, participants were guided back to the first table. Here, following the workflow shown in Figure 6, participants filled in the NASA-TLX questionnaire (Yurko et al., 2010) that measured their subjective cognitive workload during the task. After this, some open questions were asked during a short interview that had the goal of measuring what their thoughts and feelings were about the task. During this interview, it was measured whether participants were familiar with the task, whether they had strategies to remember the instructions, what influence the interruption had on their strategy and what their feelings and thoughts were about the task. After this, participants were offered a short break, together with a drink and a cookie.

After the break, the general mental capabilities of the participants were measured by using a slightly adapted version of the MMSE questionnaire from Kok and Verhey (2002a).

After this, a short Working Memory Test was administered (Stopford et al., 2012), and finally, the IADL Questionnaire from Lawton and Brody (1969) was conducted.

After this, participants were complimented on their results, debriefed and thanked for their cooperation. To minimize food waste, they were offered to bring their made sandwiches and slice of gingerbread with them.

Measurement

The specific demographic questions that were asked can be found in Appendix C.1. They include the basic questions about age and gender and whether participants had an official diagnosis of dementia. Next to this, participants were asked whether they have children, are caregivers for someone and what kind of job they previously had/have. These questions were asked because it was expected that when someone had kids or was a caregiver, they were more familiar with packing a lunch. The same holds for jobs in the catering industry. Educational level was used as a predictor as well in the data analysis, to see whether it influenced performance on the NAT.

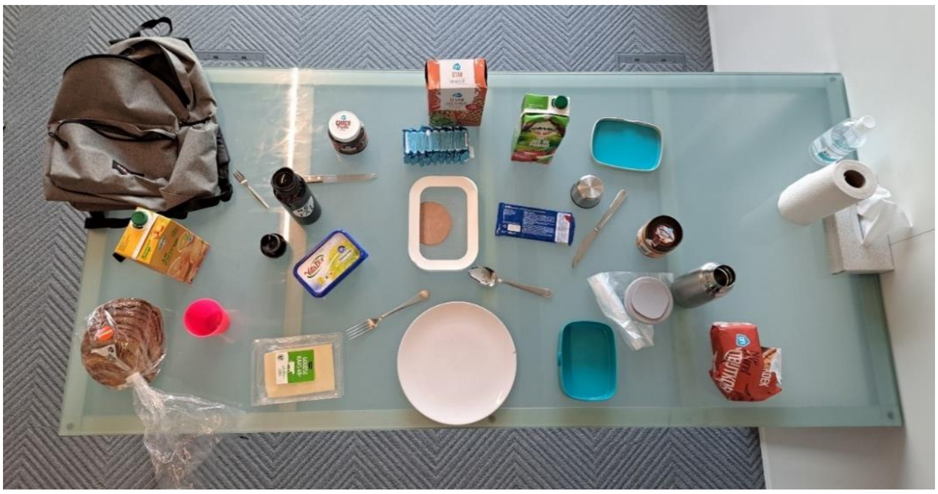
As explained, the participants had to make a packed lunch which was based on the third NAT task (Schwartz et al., 2002), except for a few changes. First of all, in the current design, all the objects were present on the table. Secondly, in the current design, we changed the order of the tasks. Participants were first asked to pour a drink and then make the sandwiches and the snack. We chose this order to create a good timing opportunity for the interruption at around one-third of the task and to ensure enough steps before and after the interruption to see performance changes caused by the interruption. Moreover, to make the task more intuitive for Dutch older adults, the original peanut butter jelly sandwich was transformed into two sandwiches, one with peanut butter and one with cheese. Finally, the cookie was changed to a slice of gingerbread, also to make the test more intuitive for Dutch older adults.

To successfully complete the three tasks, 44 steps were necessary, creating the possibility of 44 omission errors. Next to the objects needed for successful target completion, distractor items were present on the table, just as in the original NAT from Schwartz et al. (2002). These distractor items were not needed for successful task completion. Earlier research from Giovannetti et al. (2010) and Rodríguez-Bailón et al. (2017) showed that distractor items can interfere with task execution, especially when they are visually and functionally similar to the target objects or when distractor items together can also be used to create an additional contextually similar task.

The current study tried to understand how people react to an interruption in a naturalistic setting. As distractor items are usually present in naturalistic settings, they were included in this study as well. There were 13 distractor items, which also resulted in the possibility of 13 substitution errors. Both the target and distractor objects can be found in Appendix C.2. Figure 7 shows the arrangement of the objects on the table.

Figure 7

The Arrangement of the Objects on the Table



The dependent variables measured during the NAT can be found in Tables 2,3 and 4. Errors were measured per task and added together for a total score, so performance per task and performance on the whole NAT can be compared. The total score also included errors that were made before, between and after the three tasks.

Table 2 shows that not only the total amount of errors of the whole NAT was measured, but also the total amount of errors made before and after the interruption. Moreover, the definitions of Omission and Substitution Errors are specified in Table 2, together with the definitions of Tool Omissions and Task Additions.

Table 2

The Errors Measured during the ADL Task

Type of error	Definition
Total Errors	The total number of errors made
Total Errors Before the Interruption	The total number of errors made before the interruption
Total Errors After the Interruption	The total number of errors made after the interruption
Omission Error	The failure to complete a step necessary for successful task completion[1,2]
Substitution Error	The usage of a distractor object instead of a target object[2,3]
Tool Omission	Omitting a tool that is supposed to be used[2,3]
Task Addition	The execution of an extra, not task-related, step[1]

Note. Definitions based on [1]Weakley and Schmitter-Edgecombe (2019), [2]Rodríguez-Bailón et al. (2017) and [3]Giovannetti et al. (2010)

Table 3 lists the three different types of sequence errors and their definitions. As can be seen, Perseveration and Repetition Errors are closely related and are defined by the moment of the repetition of a step, either immediately (Perseveration Error) or later in the sequence (Repetition Error). A Sequence Failure is not related to the repetition of a step, but to the order of steps and whether it prevents successful task completion.

Table 3

The Different Types of Sequence Errors

Sequence errors	Definition
Perseverations	The immediate repeat of a task step
Repetitions	The repetition of a step later in the task sequence
Sequence Failures	Failures in sequence that prevent successful task completion

Note. Definitions based on Rodríguez-Bailón et al. (2017)

During the execution of the NAT, time was tracked closely. Table 4 shows all the different time-related variables measured during the execution of the NAT and how they were measured. Foremost, the time needed for task completion was measured and also divided into three parts; first, the time the participant worked on the task before the interruption. Secondly, the duration of the interruption, and finally, the amount of time necessary to complete the NAT after the interruption. To ensure equal and objective tracking of time between participants, it was chosen to use the moment that the participant arrived at the table as the starting point of the NAT and the moment the participant indicated to be finished as the ending point. To measure the duration of the interruption, the moment the participant left the table for the interruption was taken as the starting point, and the moment the participant arrived back at the table again as the ending point. Additionally, the resumption lag and the time needed to complete the three different tasks were measured as well. For Task 2, two different durations were calculated, one that includes and one that excludes the duration of the interruption.

To ensure an objective way of measuring the duration of the three tasks, it was decided that tasks started when a participant executed a step that was necessary to complete that task. For example, when the participant was supposed to start with Task 1 and opened the butter, this would not count in the duration of Task 1, as this step was not part of Task 1.

However, had the participant grabbed the steel thermos bottle first, a step that is necessary for Task 1 completion, then the grabbing of the butter would be taken into account for the duration of Task 1, as the time working on Task 1 started when the participant grabbed the steel thermos bottle.

Table 4

The Different Time-Related Dependent Measures

Dependent measure	Definition	Measurement
Total Completion Time	The time it took the participant to complete the NAT.	The number of seconds between arriving at the table until indicating to be finished. If a participant performed a step after this, that moment was taken as the ending point.
Time on Task Before the Interruption	The time the participant worked on the NAT before the interruption started.	The number of seconds between arriving at the table and leaving the table again for the interruption.
Duration of the Interruption	The time the interruption took.	The number of seconds between leaving the table for the interruption until arriving again afterward.
Time on Task After the Interruption	The time needed to complete the NAT after the interruption.	The number of seconds between arriving at the table after the interruption until indicating to be finished. When the participant performed a step after this, that moment was taken as the ending point.
Resumption Lag	The time a person needs to (re)start on the task after an interruption.[1]	The number of seconds between arriving at the table after the interruption until performing a step.
The time needed for Tasks 1, 2, and 3.	The total amount of time necessary for the participant to complete a task.	The moment that a participant started with the first appropriate step until the participant finished the last appropriate step.

Note. [1]Altmann and Trafton (2004).

Additionally to the errors described above, this study also analyzed micro errors made by participants. They were classified into three different levels based on the study by Divers et al. (2021). The first level was a Mild Micro Error, which would correspond to reaching for a wrong object. The second level was a Moderate Micro Error and consisted of reaching, and touching a wrong item and the last level was a Severe Micro Error when participants moved a wrong item (Divers et al., 2021). Moreover, micro errors were considered when participants canceled their movement or step as this indicated the participants' belief that they made an error. For all micro errors, it was also noted when they occurred, between or within a sub-task. Between errors were defined as errors that happened after the successful completion of a task and before executing the first successful step of the next task, or in the case of the first task, errors that happened before the first successful step of making the drink. Micro errors after the third task were also seen as between errors, as they did not occur within a task. Finally, whether the micro errors occurred with target objects or distractor objects was noted.

While doing the data processing, it was realized that many movements and behaviors did not fall within the predetermined errors and would be lost when doing the data analysis. That is why it was post-hoc decided to create four more variables, which are specified in Table 5. These new variables were made to enrich the data and to give some interesting insights.

The first added variable was Cleaning Addition, which involved extra steps that were performed in order to clean up the table or the objects. These steps were not necessary for successful task completion but they still added value to the data collection. It was expected that people who performed more Cleaning Additions had fewer errors as they were more aware of the situation on the table. Examples of Cleaning Additions were closing the orange juice, butter, peanut butter, cheese, and gingerbread. Next to this, it was observed that many participants cleaned their knives by sliding them over the bread slice, making the knife clean from spreads. This behavior was also noted as a Cleaning Addition. However, Cleaning Additions should be interpreted with caution, as one participant indicated in the interview that the instructions did not entail cleaning up. It is, therefore, possible that some participants would have cleaned up but did not do so as the instructions did not mention this.

The second added variable was Efficiency Step, which involves combining steps and thus creating a more efficient task. Examples of this are grabbing two slices of bread at the same time, buttering the two slices at the same time, or putting the two sandwiches at the same time in the lunch box. Moreover, there were two types of knives present at the table. One for making the sandwiches and a sharper one for slicing the gingerbread. As it was not instructed when to use which knife, the choice of the knife could not be seen as an error. However, when participants selected the sharper one for slicing the gingerbread this would be scored as an Efficiency Step. All the examples given are behaviors that were not instructed, but as these behaviors made the task easier, they were not seen as errors. It was hypothesized that people who performed these behaviors were thinking about the task themselves instead of simply following instructions, showing that they were capable of higher-order thinking. That is why it was expected that people who performed these Efficiency Steps made fewer errors in the task.

The third added variable was Organizational Additions, which consisted of added steps that had the purpose of reorganizing the objects on the table. Putting down used objects was not seen as an Organizational Addition as this movement was necessary. However, when objects were moved for the second time, for example, to make room for other objects, this would be noted as an Organizational Addition as this step was unnecessary if participants had thought earlier about where to put their objects in the first place. This is why it was expected that participants with more Organization Additions make, on average, more errors.

The last added variable was Step Recall, which is a variable that notes when participants forgot a step in the first place but later remembered it and restored their error. For example, when forgetting the cap on the thermos bottle in the first task, but only realizing and fixing this in the second task. The originally specified errors did not account for these behaviors, which is why Step Recall was created. It was expected that when people had difficulties remembering the necessary steps, they were more likely to make more errors in total as well.

As the variables Cleaning Addition, Efficiency Step, and Organizational Addition were post-hoc created and were mostly related to extra steps surrounding the tasks; they were not added to the Total Amount of Errors variable. On the contrary, as Step Recall only involved steps needed for successful task completion, this variable was included in the Total Amount of Errors.

Table 5

The Post-Hoc Created Variables

Variable	Definition
Cleaning Addition	Cleaning up the table after the usage of objects.
Efficiency Step	Combining two steps or choosing the better option between target objects.
Organizational Addition	Re-organizing the objects on the table without using the objects.
Step Recall	Execution of a step later in the sequence as it was forgotten earlier on.

As explained, an interruption occurred within the NAT. This interruption consisted of four questions, which are placed in Appendix C.3. They were made to target different aspects of memory. For example, the question "*What did you have for breakfast this morning?*" targets retrospective memory, whereas the question "*What are you going to eat for dinner tonight?"* targets prospective memory. Moreover, it was asked how the participant got to the campus. This question was asked because it was expected that it would require more cognitive workload than the other questions as this trip to the campus is not a regular one.

The NASA-TLX was administered using six items to measure the subjective cognitive workload of the participants. The original version was made by Hart and Staveland (1988), but the version from Yurko et al. (2010) was used to translate the NASA-TLX to Dutch. The translated version was checked by an independent researcher and also compared with the Dutch version made by Bemer (2019). After finalizing the translation, the questionnaire was translated back into English by two other independent researchers and this was compared with the original English version by another independent researcher. The final result can be found in Appendix C.4. The NASA-TLX used six items, each measuring a subjective component with a single question, that were equally weighted and combined to form the Total Subjective Workload ($\alpha = 0.72$, according to Hoonakker et al. (2011)). The average Subjective Workload could range from 0 to 100 and scores were interpreted as "low (0-9), medium (10-29), rather high (30-49), high (50-79) and very high (80-100)" (Sugarindra et al., 2017, p.4). It should be noted that the named interpretation scores are normally used for weighted NASA-TLX scores. However, as no other interpretation scores were known to the authors, it was chosen to use the one explained above.

Items could also be used individually to see the effect of one subset of cognitive workload. The items measured by the NASA-TLX were Mental, Physical, and Temporal Demand, Performance, Effort, and Frustration (Hart & Staveland, 1988). All individual items were scored on the same scale, except for Performance, as for this item, a high score meant high confidence in successfully executing the instructions, and thus a low subjective workload. That is why for the data analysis, Performance was reverse coded. Examples of questions were: "*How frustrated were you during the task?"* to measure Frustration and *"How rushed did you feel during the execution of the task?"* to measure Temporal Demand

(Yurko et al., 2010). Participants could answer on a range from 1 (bijna niet; very little) to 100 (heel erg; a lot), with steps of 5.

The interview was conducted according to a semi-structured format. The questions used for the interview can be found in Appendix C.5. Example questions were: *"Did you have a certain strategy to remember the instructions?"* and "*Were the actions you had to perform familiar to you?".* After the interview, the audio files were transcribed. The transcription was based on the system from Bailey (2008) that, for example, uses two dots to indicate a silence of less than one second. After transcription, the text was manually coded by the experimenter.

The MMSE was used to measure the cognitive functioning of the participants. The original version was created by Folstein et al. (1975). In this research, the Dutch version from Kok and Verhey (2002a) was used. In this 11-item questionnaire, a total of 30 points can be scored. The score was used as an indication of the cognitive functioning of participants as a score lower than 26 is generally seen as a sign of mild cognitive impairment (Zadikoff et al., 2008). Item 2 of the MMSE was slightly adjusted as the experiment did not take place in a hospital. Based on Molloy and Standish (1997), item two now asked participants to name the country, province, city, and building they are currently in. Next to this, participants had to mention what floor they were on. The scoring used was explained in Kok and Verhey (2002b) and was based on the papers from Folstein et al. (1975) and Molloy et al. (1991). Example tasks of the MMSE are to name a pencil and a watch and to repeat the sentence: "*Nu eens dit en dan weer dat".* The full MMSE with the amount of points per item can be found in Appendix C.6.

Working memory was tested by a word reading and recalling task. The original task is explained in Stopford et al. (2007) but in this research, the shorter version from Stopford et al. (2012) was used. In each trial, participants received a paper with three monosyllabic and semantically unrelated words (Stopford et al., 2012). To minimize possible confounding effects, each trial consisted of three words that were phonologically and visually different from the words in the previous trial (Stopford et al., 2007). Participants were instructed to read the words out loud, but also to remember them. Three conditions were tested, the first involved immediate recall of the three words. The second condition involved a five-second delay before recall and the last condition required the participants to read a sequence of numbers during the five-second delay before recall (Stopford et al., 2012). Each condition had six trials, split into two groups that were administered in counterbalanced order. The score can range from zero words recalled to 54 words recalled, the more words recalled the better the working memory The words and numbers used in the working memory test can be found in Appendix C.7.

A self-made translation of the IADL Questionnaire from Lawton and Brody (1969) was used to measure the general IADL independence of the participants. After the questionnaire was translated, three independent researchers translated it back into English. Two more independent researchers judged the differences between the original English version and the two translated versions. Some minor alterations were made, and the final Dutch IADL Questionnaire can be found in Appendix C.8. The IADL Questionnaire tests independence in eight categories, for example, household. The experimenter asked participants to explain how they deal with each category and, based on this, rated the independence of this category. Scores per category range from zero (dependent) to one (independent). Therefore, the maximum score of eight will suggest that the participant is fully independent.

Data analysis

Quantitative analysis

After the data collection, the video footage was reviewed, and all the steps performed by the participants were written down. Next, the data was analyzed twice, once to code all the steps and a second time to make sure that when similar situations occurred between different participants, they were coded the same. To limit subjectivity, the coding was checked independently with another researcher. After the coding was complete, a scoring form for all participants was filled in, in which all the errors were counted. This form was then transferred to Excel and combined with the other data collected within the study. This Excel form was then loaded and analyzed in STATA (StataCorp, 2019). In the case that participants did not want to continue on a task, missing values were noted down.

The next step was to inspect all the data and their distributions. This was done by looking through the data, making graphs and tables and checking for outliers. The identification of outliers happened by standardizing the variables and checking for z-scores outside of the -3 to 3 range. When outliers were present, the model was run twice, once including the outliers and once excluding them. As the identified outliers were no measurement errors, it was decided to report the model with the original data in the result section. In the case that the removal of outliers made a significant difference to the outcome of the model, both the model with and without outliers would be reported. A significant difference was defined as a change in significance for a predictor or a change of sign for a significant predictor.

Multilevel regression was used to test the relationship between the variables. Predictors used were age, gender, number of children, level of education, and the scores on the MMSE, Working Memory Test, NASA-TLX, and the IADL Questionnaire. The outcome variables were the dependent measures collected by the NAT, for example, Total Completion Time and Total Number of Errors. Moreover, the speed-accuracy trade-off was measured by using the Total Number of Errors as the outcome variable and Total Completion Time as the predictor. The speed-accuracy trade-off was also tested before and after the interruption by using the Total Amount of Errors before and after the interruption as the predictors for the outcome variables Time on Task Before the Interruption and Time on Task After the Interruption. Furthermore, the different types of errors were used as predictors for Total Completion Time to see whether or not different types of errors impacted the duration of the task.

For all multilevel regression models, the normality of the residuals was checked with the Shapiro-Wilk and skewness/kurtosis test. If the normality of residuals was violated, it was first tried whether the removal of outliers helped. If this was not successful, input variables were transformed. Additionally, multicollinearity was checked for each model by using the "estat vif" command in STATA. When the individual VIF items were below 8 and the mean VIF value was below 2.5, it was concluded that this assumption was met. Heteroscedasticity was checked as well, by using the "imtest", "hettest" and "hettest, rhs". When these tests resulted in a nonsignificant number, the conclusion was drawn that heteroscedasticity was not a problem, otherwise, a robust multilevel regression will be performed.

As multiple regressions were performed on the same data, a correction for multiple testing was needed. However, with the amount of regressions performed (18), the Bonferroni and Bonferroni-Holm correction appeared to be too conservative, for example, the Bonferroni corrected alpha became 0.003 (0.05/18). It was concluded that while guarding against multiple testing is important, it should not lead to too many missed findings. The current research used three main outcome variables, MMSE score, Total Completion Time, and Total Amount of Errors. Other outcome variables related to those three variables, for example, the

Duration of the Interruption is a part of the Total Completion Time. As all outcome variables could be retraced to three variables, it was decided to divide the normal alpha value of 0.05 by three (0.017) and use this value as alpha for the current research. Although unconventional, it was believed that the current approach was the best for the current research.

To make sure the two different orders of the Working Memory Tests, did not influence the outcome, two T-tests were performed. The first compared the means of the total amount of words recalled from the full Working Memory Tests and the second only compared the mean of the total amount of words recalled from the third and hardest trial. For both T-tests, whether the variables were normally distributed and whether their standard deviation was the same was checked.

Qualitative analysis

After all interviews were conducted, they were transcribed and coded based on the topic of the quote. A second round of coding was then conducted to create codes within the topics. Afterward, all the codes and topics were analyzed, and themes and sub-themes were identified. For this thematic analysis, several choices were made based on the paper from Braun & Clarke (2006). The first decision was about the classification of themes, and it was decided that a patterned response would transform into a theme if mentioned more than 4 times by different participants. However, the researcher also could make a theme of a response that occurred less than 4 times, if deemed important. As the interview was used for exploratory reasons, the whole data set was used, and the thematic analysis was inductive, meaning data-driven instead of theory-driven, which were the answers to the second and third choices proposed by Braun & Clarke (2006). Moreover, both semantic and latent themes were used (Braun & Clarke, 2006). This was because latent themes look more at the meaning behind the answers that were given and could add crucial information to the outcome of the study. Lastly, the thematic analysis used the constructionist view (Braun & Clarke, 2006) as it was used to create theories that could be generalized to the whole older adult population.

Pilot

To make sure that the methodology was working properly and to verify the assumptions made, a pilot was held. This was done at "de Meerpaal", a community center in Eindhoven on two afternoons. The pilot consisted of two participants, and it was decided not to collect more as participant recruitment was expected to be hard. Based on the pilot, several changes to the procedure were made, mostly related to the introduction of a second table. During the pilot, it was noticed that the usage of one table had a few limitations. First, the table was full of objects, making it harder to administer the questionnaires and perform the Working Memory Test. Secondly, when taking place at the table, participants immediately looked, grabbed, and moved the objects for the NAT. This behavior was even stronger when explaining the task. The intention was that the participants first hear the whole instruction and then see the objects for the first time. To achieve this, it was decided after the pilot that participants in the real study would first sit at a different table and get the instructions there.

The pilot also made clear that during the interruption, participants also needed to sit at another table. This is because, during the interruption, the participant still tried to work on the NAT, which made the interruption less effective. During the second pilot, this was tested, and it was noticed that the participant paid more attention to the interruption. Lastly, the lunchbox and the sandwich bags were replaced with easier closable ones as participants struggled with the opening mechanisms.

Results

The total number of participants was 21, of which only one person had an official diagnosis of dementia. As this disease is expected to influence the results of the ADL task and the questionnaires, it was decided to keep the data from this person as exploratory data and not combine it with the data of the other 20 participants. The data from the person with dementia can give some insight into the differences between the two groups. However, as the second group only consisted of one person, no real conclusions can be drawn about the differences between the two groups.

Descriptive Analysis

Subjective Workload

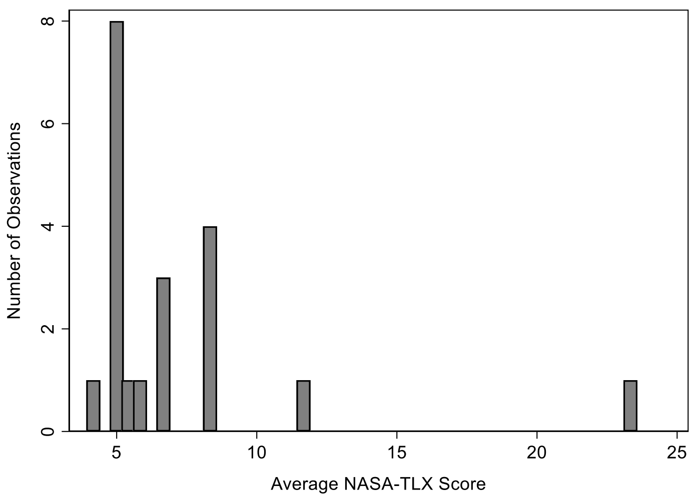
The mean of the average Subjective Workload measured by the NASA-TLX was 7.19 (4.17-23.33, *SD:* 4.23*)*, indicating that participants experienced low mental workload (Sugarindra et al., 2017). Figure 8 shows the distribution of the average NASA-TXL scores, and two obtained scores stand out by being far from the others. Indeed, the standardization of the average NASA-TLX scores showed that the obtained score of 23.3 could be a possible outlier. In the interview, this participant noted that the task itself was not very difficult, but more the location, explaining his high score. Based on this, it was decided to regard this observation as an outlier.

The skewness value for the total NASA-TLX score was 3.00 which falls outside the -2 to 2 range, indicating the possibility of rejection of normality. However, various thresholds are used (Watson, 2018). The same holds for Kurtosis, as the measured kurtosis value was 11.93, which falls outside of the -7 and 7 range. The Shapiro-Wilk and skewness/kurtosis test indeed rejected normality (p <.001 for both tests). Removing the outlier did not result in a different outcome. Detailed information about the individual items of the NASA-TLX score can be found in Appendix D.1.

The person with dementia had an average score of 19.7, which is higher than the average score of the participants without a dementia diagnosis but not the highest score.

Figure 8

The Distribution of Average NASA-TLX Scores



Cognitive Workload

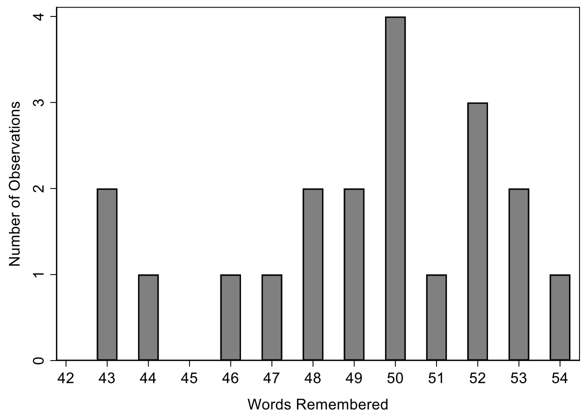
MMSE. The mean MMSE score (28.45) approximates the perfect score, being 30 (*SD:* 1.05, 26-30). However, only 3 participants did earn the full 30 points. Most participants earned 29 (7 participants), or 28 (also 7 participants) points. 2 participants earned 27 points and only one participant earned 26 points. The skewness (-0.42) and kurtosis (2.88) values for the total MMSE score fall within the -2 and 2 range for skewness and the -7 to 7 range for kurtosis. Indeed, both the Skewness/kurtosis and Shapiro-Wilk test showed that normality was not rejected. No outliers were detected. Detailed results about the individual items of the MMSE can be found in Appendix D.2.

Working Memory Test. Figure 9 shows the distribution of the scores obtained in the Working Memory Test. The mean score was 49.3 (*SD:* 3.27) and the scores ranged between

43 and 54. The skewness (-0.57) and kurtosis (2.40) values fell within the earlier discussed ranges and indeed the Skewness/kurtosis and Shapiro-Wilk test did not reject normality. Moreover, no outliers were found. Detailed results of the three trials of the Working Memory Test can be found in Appendix D.3.

Figure 9

The Distribution of Working Trial Scores



The person with dementia completed the first two trials without any error but scored lower in the final trial as he remembered only 9 words, giving him a total score of 45.

Daily IADL Performance

The IADL Questionnaire consisted of eight categories. One of those categories was about being responsible for one’s own mediation. However, not all participants had medication, in which case no points could be earned or lost in that category. To make up for this, it was decided to calculate a percentage score for each participant on the IADL Questionnaire, compensating for the number of applicable categories and the points scored in them. This meant that a person that does not take medication, could still earn a 100% score as the medication category was removed for that person. Table 6 shows the scores and their distributions for the IADL Questionnaire. As can be seen, for most participants (14) all eight categories applied. Only for 6 participants did one category not apply of which 5 still got a full score. The mean IADL score was 0.95 (*SD:* 0.08) and the skewness and kurtosis values were -1.11 and 3.15, respectively. The Skewness/kurtosis test did not reject normality and no outliers were found, as well. Still, the Shapiro-Wilk test rejected normality (p = 0.002).

Table 6

The Scores and Distributions on the IADL Questionnaire

IADL Score (%)	Number of Participants	Number of Points	Number of Categories
75%	1	6	8
85,7%	1	6	7
87.5%	5	7	8
100	5	7	7
100	8	8	8

The person with dementia scored 0% on the IADL Questionnaire which can be explained by the fact that he lives in a residential care center.

Duration of the NAT

Table 7 shows all the variables regarding the duration of the NAT. Participants took around 6 minutes to complete the full NAT, including the duration of the interruption. Task 2 took the longest to complete (*M:* 134.5, *SD:* 40.31), compared to Task 1 (*M:* 40.8, *SD:* 7.96) and Task 3 (*M:* 30, *SD:* 12.59). The mean Duration of the Interruption was somewhat more than two minutes, which is close to the duration of Task 2. The interruption took place around one-third of the task when looking at the mean Time on Task Before (76.25) and After (169.8) the Interruption.

Table 7

The Durations of the Task and Sub-Tasks in Seconds

Variable	Observations	Mean	SD	Min	Max	Skewness	Kurtosis
Total Completion Time	20	371.95	65.07	265	528	0.51	2.99
Total Completion Time Excluding the Duration of the Interruption	20	247.05	54.14	171	396	1.34	4.60
Task 1 Completion Time	20	40.8	7.96	29	59	0.83	3.21
Task 2 Completion Time	20	261.35	52.49	186	360	0.23	1.94
Task 2 Completion Time Excluding the Duration of the Interruption	20	134.5	40.31	86	228	0.93	3.27
Task 3 Completion Time	20	30	12.59	86	228	0.01	2.34
Time on Task Before the Interruption	20	76.25	15.26	61	127	1.83	7.08
Time on Task After the Interruption	20	169.8	48.37	108	305	1.42	4.82
Duration of the Interruption	20	125.9	35.22	80	238	1.49	6.13
Resumption Lag	20	0.85	0.67	0	3	1.24	6.56

For all variables, the skewness values were between -2 and 2, and only for one variable the Kurtosis value was outside the -7 to 7 range as Time on Task Before the

Interruption had a Kurtosis value of 7.08. However, the skewness/kurtosis test showed that normality was rejected for the Total Completion Time Excluding the Interruption ($p = 0.012$), Time on Task Before the Interruption ($p = 0.001$), the Duration of the Interruption ($p = 0.003$) and the Time on Task After the Interruption ($p = 0.008$). The Shapiro-Wilk test rejected the normality for the same variables but also rejected normality for the variable measuring the duration of Task 2 excluding the duration of the interruption ($p = 0.047$).

In each of the variables Time on Task Before the Interruption, the Duration of the Interruption, and the Resumption Lag, one possible outlier was detected. When removing all three observations, all variables passed the Shapiro-Wilk and the Skewness/Kurtosis test. The one outlier that was detected in the Resumption Lag variable was the only duration of three seconds. Other observations were a zero-second resumption lag (5 participants) and a one-second resumption lag (14 participants).

The person with dementia took longer on all three tasks: 68 seconds for Task 1, 280 seconds for Task 2, and 75 seconds for Task 3. Also, the Total Completion Time was the longest of all participants (592 seconds). The interruption lasted 100 seconds, which is comparable to the duration of the interruption of the healthy older adult group. When removing the duration of the interruption from the Total Completion Time, the resulting time (492 seconds) was substantially larger compared to the other participants. The participant with dementia worked for 133 seconds on the task before the interruption started, which is already longer than the healthy participants did. Moreover, the Time on Task After the Interruption (359) was slightly longer than the maximum value seen in the group of participants without an official diagnosis of dementia.

Errors on the NAT

When looking at Table 8, it can be found that in total 220 errors were made in the NAT (*M:* 11, *SD:* 4.86), the number of errors made per participant ranged between 2 and 21. Most of the 220 errors occurred after the interruption (137 errors or 62%). Interestingly, no participant made a Tool Omission Error and only one participant made a Substitution Error. Moreover, there were only two participants who made a Perseveration Error. The most common error was a Task Addition, as in total 60 Task Additions were registered (*M:* 3, *SD:* 2.77*)*. The second most common error type was Omission Errors, with a total of 56. The amount of Mild, Moderate, and Severe Micro Errors did not differ significantly. Interesting to note is that there were more Severe Micro Errors than Mild and Moderate Micro Errors.

Table 8

The Total Amount of Errors Made

Variable	Total Amount of Errors	Observations	Mean	SD	Min	Max
Omission	56	20	2.8	2.12	0	6
Substitution	1	20	0.05	0.22	0	1
Tool Omission	0	20	0	0	0	0
Perseveration	3	20	0.15	0.49	0	2
Repetitions	36	20	1.8	1.40	0	6
Sequence Failure	16	20	0.8	1.44	0	6
Step Recall	13	20	0.65	1.57	0	6
Task Addition	60	20	3	2.77	0	12
Mild Micro Error	12	20	0.6	0.68	0	2
Moderate Micro Error	9	20	0.45	0.67	0	2
Severe Micro Error	14	20	0.7	0.73	0	2
Combined	220	20	11	4.86	2	21

Figure 10 shows the error distribution per participant. The Shapiro-Wilk and skewness/kurtosis test did not reject normality for the Total Amount of Errors, and no outliers were detected.

Figure 10

The Distribution of the Total Amount of Errors Made

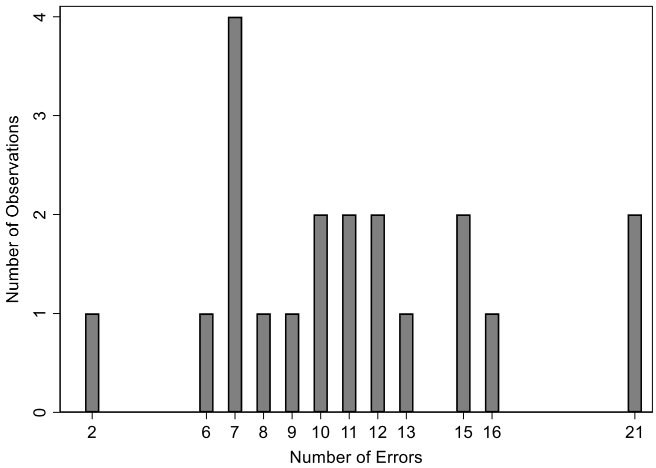


Figure 11 shows the distribution of five different errors. It can be seen that most Omission Errors happened in Task 1. Further analysis showed that they were almost all related to the participants forgetting to put the lid or cap on the thermos bottle. Three participants forgot to put the steel thermos bottle and lunch box in the backpack, explaining the high amount of Omission Errors after Task 3. Repetitions were mostly made in Task 2 and were because of making the sandwiches. One person started the NAT with the second task, creating some Sequence Failures at the beginning. Sequence Failures were also common in Task 1 as participants had difficulties with the order of the lid and cap on the steel thermos bottle. One person indicated to be finished with the NAT and before walking to the other table, remembered that the thermos bottle and lunch box needed to be packed in a backpack, these Step Recalls are indicated with the right bar in Figure 11. Task Additions were common across all three tasks. Only one Substitution Error was made, as one participant switched the orange juice for the apple juice in Task 1. Three Perseveration Errors were made, two in Task

2 and one in Task 3. For readability reasons, both the Substitution and Perseveration Errors were not shown in Figure 11. Appendix D.4 shows more detailed information about the individual error variables.

Figure 11

The Composition of the Errors across the NAT

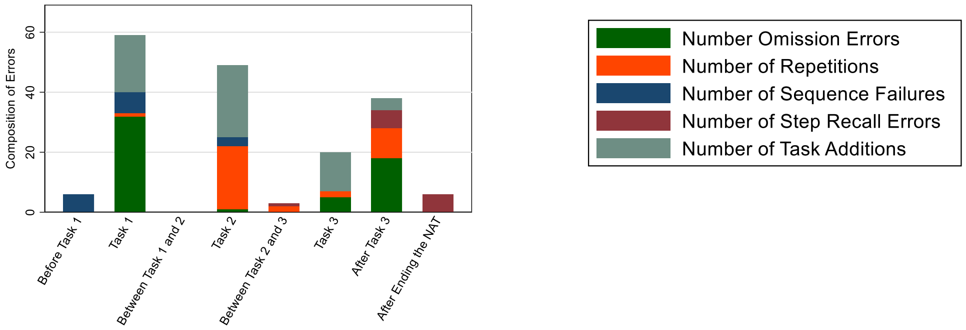
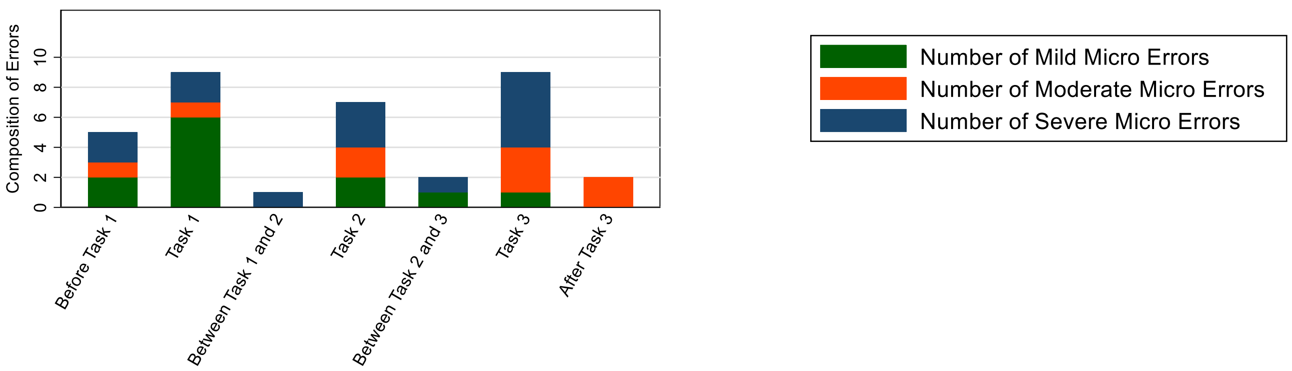


Figure 12 shows the three different types of micro errors and their occurrences during the NAT. Most micro errors were made within the three tasks or before starting the first task. Noteworthy is that Mild Micro Errors were predominantly made at the beginning of the NAT, whereas Moderate Micro Errors occurred at each moment during the NAT. Severe Micro Errors happened mostly during the last task. Appendix D.5 shows more detailed information about the micro errors made during the NAT.

Figure 12

The Composition of Micro Errors across the NAT



When looking at Table 9, it can be seen that every participant performed some sort of cleaning behavior, as the minimum value of Cleaning Addition is 3 (*M*: 6.6, *SD:* 1.96). This also explains the large number of occurrences (132). Neither Efficiency Steps nor Organization Additions were executed by all participants. Efficiency Steps in general were not often performed, it only occurred 25 times, while the maximum number of occurrences is 4 as well. Organizational Additions were more common, with a mean of 3.4. However, here, the standard deviation is quite large (2.58).

Table 9

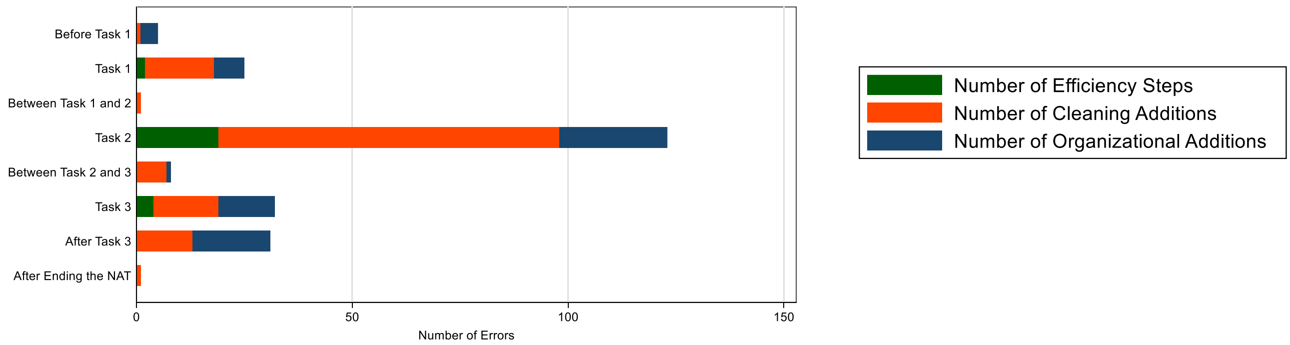
The Occurrences of the Post-Hoc Created Data Enrichment Variables

Variable	Amount of Occurrences	Observations	Mean	SD	Min	Max
Efficiency Step	25	20	1.15	1.41	0	4
Cleaning Addition	132	20	6.6	1.96	3	9
Organizational Addition	86	20	3.4	2.58	0	8

Figure 13 shows how many Efficiency Steps, Cleaning Additions, and Organizational Additions were made and at what moment during the NAT. As explained, most Efficiency Steps could be made during Task 2, which is also shown in Figure 13. Cleaning Additions were mostly made during a task and predominantly during Task 2. This was expected as during Task 2 most objects were necessary which created the possibility of Cleaning Additions. Remarkable is the fact that one participant left the table after finishing the NAT and came back to clean the table some more. Figure 13 also shows that participants sometimes waited until finishing tasks before cleaning. Although some Organizational Additions were executed when arriving and starting the NAT, most occurred later during the NAT, in Tasks 2 and 3. More detailed information about the data enrichment variables can be found in Appendix D.6.

Figure 13

The distribution of the Data Enrichment Variables



Outliers. In six of the discussed error variables, possible outliers were detected. As mentioned before, only one person made a Substitution Error and only two people made a Perseveration Error, of which one participant made two of them. When standardizing those two variables, the Substitution Error and the two Perseveration Errors were indicated as possible outliers. However, one Substitution Error and the two Perseveration Errors were not judged to be real outliers, also because they were no measurement errors. That is why they were not excluded from the data set.

Additionally, one person made four instead of two sandwiches, resulting in six Repetitions. Furthermore, one participant started the NAT by working on the second task, resulting in six Sequence Failures. Another participant almost forgot to put the steel thermos bottle and the lunch box in the backpack, resulting in six Step Recalls. Finally, one participant packed the sandwiches in plastic sandwich bags, resulting in eight Task Additions. Standardization showed that these four observations could be seen as outliers and could, thus, possibly affect the results of the hypothesis testing. That is why models including these variables were tested twice, once with and once without the outliers.

Errors Made by the Participant with Dementia

The person with dementia had a total of 28 errors, which is substantially higher than the number of errors made by participants without a diagnosis of dementia.16 of the 28 errors occurred after the interruption. Interestingly, the participant with dementia did make two Tool Omission Errors as he did not use the knife to slice the bread, instead, he folded the bread twice. Most of the errors were Task Additions (12). However, they mainly occurred simultaneously. This was because the participant had trouble pouring in the orange juice. He started by putting the cap on the thermos bottle (his only Sequence Failure). Next, he removed the gap and grabbed the orange juice, opened it a bit, and held it upside down spilling some orange juice on the floor. Afterward, he poured orange juice into the cap and used the cap to pour the orange juice into the thermos bottle. This sequence of steps resulted in six Task Additions, and as he forgot the lid on the thermos bottle, also resulted in three Omission Errors.

Other findings were that this participant moved a lot of objects around the table, resulting in more Severe Micro Errors (6), Organizational Additions (9), and Repetitions (3). Noteworthy, is the fact that this participant also performed two Efficiency Steps as he grabbed and buttered two bread slices at the same time. Moreover, this participant performed seven Cleaning Additions. As mentioned, this participant made six Severe Micro Errors, which is significantly more than the healthy older adults. It is also worthwhile mentioning that this participant did not make a Mild Micro Error and only one Moderate Micro Error.

Hypothesis testing

Before performing the regression analyses, it was first tested whether counterbalancing the order of trials in the Working Memory Test had a significant impact on the mean reproduction score. This was both done for the whole Working Memory Test and for the third trial, as that trial had the most variability. A T-test showed that the mean reproduction score for both the whole Working Memory Test (t(18)=0.67, *p*=0.509) and trial three (t(18)= 0.57, *p* = 0.575) did not differ between the two different orders, indicating that counterbalancing the order did not impact the outcome on the Working Memory Test.

Starting with the regression analysis, it was first tested whether demographic conditions influenced the total MMSE score. Table 10 shows the outcome of this regression. All variables combined did not significantly predict the total MMSE score (F(5,13) = 0.28, *p* = 0.914, $R^2_{Adjusted}$ = -0.249). Moreover, all predictors were individually also not significantly related to the total MMSE score.

TABLE 10

Results of the Multilevel Regression with MMSE Score as Outcome Variable

Variable	B	SE	t	p-value	95% CI
Age	-.0328	.0627	-0.52	0.609	-0.168:0.103
Gender	.5879	.6760	0.87	0.400	-0.873:2.048
Number of Kids	-.0514	.3962	-0.13	0.899	-0.907:0.804
Education Level	.0061	.3340	0.02	0.986	-0.716:0.728
Recruitment Manner	.1970	.7280	0.27	0.791	-1.376:1.700

Note. For gender, females were the reference category.

It was also tested whether the answers on the NASA-TLX, Working Memory Test, and IADL Questionnaire were related to the outcome of the MMSE score. The outcome of this regression can be found in Table 11 and again, no significant relation was found (F(3,16) = 1.27), *p*= 0.320, $R^2_{Adjusted}$ = 0.04). Excluding the outlier from the NASA-TLX score did change the coefficient from positive to negative. However, the p-value remained above 0.05, creating a non-significant result. Remarkable is that when excluding the outlier from NASA-TLX, also the coefficient from the IADL Questionnaire became negative, although staying non-significant. Changing the scores of the Total Working Memory test for only the third trial did not cause any significant changes.

Table 11

Results of the Multilevel Regression with MMSE Score as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
NASA-TLX	0.067	0.067	0.90	0.381	-0.081:0.202
Working Memory Test	0.1471	0.080	1.85	0.083	-0.021:0.316
IADL Score	1.791	3.472	0.52	0.613	-5.570:9.152

Furthermore, it was tested whether the age and gender of the participants or their scores on the MMSE, IADL Questionnaire, NASA-TLX, and Working Memory Test influenced the Total Completion Time of the NAT. This model is shown in Table 12 below. The model itself did not have predictive power (F(6,13) = 2.12), p = 0.120, $R^2_{Adjusted}$ = 0.26). However, when looking at the individual variables, Age normally would have a significant relationship with the Total Completion Time of the NAT. However, as this research used a corrected alpha of 0.17, this relationship was not seen as significant. Still, it is interesting to see that males, in general, took almost 40 seconds more to complete the NAT. Removing the outlier on the NASA-TLX score did not result in significant changes for that variable. However, it did increase the p-value for Age to 0.072. The coefficient changed only slightly from 6.161 including the outlier to 6.479 without the outlier. However, as Age was already considered nonsignificant, it was concluded that the removal of the outlier did not result in significant changes.

Table 12

Results of the Multilevel Regression with the Total Completion Time as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	22.330	14.576	1.53	0.149	-9.158:53.819
NASA-TLX	-7.730	4.205	-1.84	0.089	-16.816:1.354
Working Memory Test	-5.064	5.268	-0.96	0.354	-16.445:6.317
IADL Score	52.825	222.438	0.24	0.816	-427.722:533.373
Gender	39.965	35.384	1.13	0.279	-36.478:116.408
Age	6.161	2.752	2.24	0.043	0.216:12.105

Note. For gender, females were the reference category.

The same six predictor variables were also placed in a model with the Total Completion Time Excluding the Duration of the Interruption as the outcome variable. This regression, which is shown in Table 13 below, has predictive power (F(6,13) = 5.01, p = 0.007, $R^2_{Adjusted}$ = 0.56) and shows that the gender of participants did influence the duration of the NAT when excluding the duration of the interruption (η_p^2 = .414). Table 13 shows that males took, in general, almost 70 seconds more to complete the NAT compared to females. With the corrected alpha of 0.17, the relationship between Age and Total Completion Time Excluding the Duration of the Interruption was not seen as significant. Removing the outlier from the NASA-TLX increased the p-value from Age from 0.035 to 0.062. However, because of the corrected alpha of 0.17, this does not matter for interpretation.

Total Completion Time Excluding the Duration of the Interruption can be split into two parts, Time on Task Before and Time on Task After the Interruption. At this moment, it was only tested whether the total MMSE score related to the Total Completion Time Excluding the Duration of the Interruption. However, further analysis was done to see whether this relation changed when using Time on Task Before or Time on Task After the Interruption as the outcome variable. This appeared not to be the case. The output of these two regressions can be found in Appendix E.1.

Table 13

Results of the Multilevel Regression with the Total Completion Time Excluding the Duration of the Interruption as the outcome variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	16.740	9.379	1.78	0.098	-3.522:37:002
NASA-TLX	-5.613	2.706	-2.07	0.058	-11.459:0.233
Working Memory Test	-4.072	3.390	-1.20	0.251	-11:395:3.252
IADL Score	176.449	143.129	1.23	0.239	-132.763:485.660
Gender	68.937	22.768	3.03	0.010	19.749:118.125
Age	4.162	1.771	2.35	0.035	0.337:7.988

Note. For gender, females were the reference category.

It was also investigated whether the same predictor variables had influenced the Duration of the Interruption. At first, the normality of residuals assumption was rejected. However, after removing the outliers from the Duration of the Interruption and the NASA-TLX, this assumption was passed. The outcome of this regression without the outliers can be found in Table 14 and shows that none of the variables were significantly related to the Duration of the Interruption ($F(6,11) = 1.02$, $p = 0.462$, $R^2_{Adjusted} = 0.01$).

Table 14

Results of the Multilevel Regression with the Duration of the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	3.796	6.458	0.59	0.569	-10.418:18.011
NASA-TLX	-3.969	4.763	-0.83	0.422	-14.452:6.514
Working Memory Test	-2.859	2.309	-1.24	0.241	-7.941:2.223
IADL Score	-147.756	104.966	-1.41	0.187	-378.786:83.273
Gender	-24.165	15.376	-1.57	0.144	-58.007:9.677
Age	2.911	1.370	2.12	0.057	-0.104:5.925

Note. For gender, females were the reference category.

The amount of errors a participant made is one of the main variables of interest. Table 15 shows the model that analyzed whether the scores on the MMSE, NASA-TLX, IADL Questionnaire, and Working Memory Test were significantly related to the number of errors participants made in the NAT. Moreover, the variables Gender, Age, and Total Completion Time were added, together with the three post-hoc created data enrichment variables. These data enrichment variables were not taken into account in the Total Number of Errors but might still relate to it. Table 15 shows the outcome of this regression (F(10,9) = 0.70, p = 0.708, $R^2_{Adjusted}$ = -0.19) which shows no relationship between the predictor variables and the Total Number of Errors made. Removing the outliers from the data set did not result in significant changes. The individual VIF values for Gender (3.91) and Age (3.74) and the average VIF value (2.60) were a little high, but not significant enough to take action. Further analysis also showed that the total MMSE score was not related to the number of errors participants made before and after the interruption. The output of these two regressions can be found in Appendix E.1.

Table 15

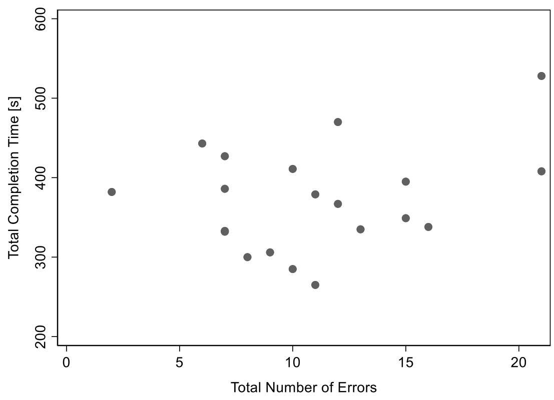
Results of the Regression with the Total Number of Errors as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	1.571	1.616	0.97	0.356	-2.085:5.227
NASA-TLX	-0.335	0.467	-0.72	0.491	-1.390:0.721
Working Memory Test	0.239	0.537	0.45	0.666	-0.976.1.455
IADL Score	-16.726	26.530	-0.63	0.544	-76.742:43.290
Gender	-3.229	4.703	-0.69	0.510	-13.869:7.410
Age	0.216	0.388	0.56	0.592	-0.661:1.093
Total Completion Time	0.024	0.032	0.75	0.470	-0.048:0.096
Cleaning Addition	-0.621	0.764	-0.81	0.438	-2.350:1.109
Efficiency Step	1.633	1.277	1.28	0.233	-1.255:4.522
Organizational Addition	0.276	0.710	0.39	0.706	-1.329:1.882

The model in Figure 15 showed that there is no relationship between the total duration of the NAT and the total number of errors participants made, rejecting a possible speed-accuracy trade-off. Figure 14 below shows the scatterplot of the Total Completion Time against the Total Amount of Errors. Figure 14 does indeed not show a speed-accuracy trade-off, instead, the data suggests that the longer participants took to complete the NAT, the more errors they made. It was also analyzed whether a speed-accuracy trade-off was present before or after the interruption. However, regression analysis rejected this. The output of these regressions can be found in Appendix E.1.

Figure 14

Scatterplot of the Total Completion Time against the Total Number of errors made



At first, also the variables Total Completion Time Excluding the Duration of the Interruption and the Duration of the Interruption were included in the model displayed in Figure 15 but they resulted in an average VIF value of 61.65, indicating multicollinearity. That is why it was decided to place them in a separate regression model, which is shown in Figure 16. This model passed the multicollinearity assumption. Whereas the Total Completion Time Excluding the Duration had a p-value of 0.044, it was not regarded as significant as the corrected alpha was set at 0.17. Removing the outlier from the variable measuring the duration of the interruption changed the p-value of the Total Completion Time Excluding the Duration from 0.044 to 0.070. However, because of the corrected alpha this did not make a significant difference.

Table 16

Results of the Multiple Regression with the Total Amount of Errors as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
Duration of the Interruption	-0.021	0.029	-0.73	0.477	-0.083:0.040
Total Completion Time Excluding the Duration of the Interruption	0.041	0.019	2.17	0.044	0.001:0.081

It was also analyzed whether the individual error variables were related to the Total Completion Time of the NAT. This regression can be found in Table 17 ($F(13,6) = 6.01$). It can be seen that the model explains a large portion of the variance of the Total Completion Time ($R^2_{Adjusted} = 0.77$) and has predictive power ($p = 0.019$). Some variables are highly significant, like the total amount of Perseveration Errors ($p = 0.006$, $\eta_p^2 = .747$), the amount of Step Recalls ($p = 0.008$, $\eta_p^2 = .718$), and the amount of Mild Micro Errors ($p = 0.012$, $\eta_p^2 = .676$).

The coefficients from these variables are positive indicating that making more of these errors was significantly related to a longer Total Completion Time. For example, making one more Mild Micro Error is related to an increase of almost 60 seconds in Total Completion Time. Moderate and Severe Micro Errors did not relate to the total duration of the task. With the corrected alpha, Substitution errors were not seen as significantly related to Total Completion Time. Task Additions, on the other hand, were just significantly related to the total duration of the NAT ($p = 0.017$, $\eta_p^2 = .639$). Although Perseveration Errors were significantly related to the Total Completion Time, conclusions should be handled with care as only two people made a Perseveration Error.

Table 17

The Results of the Regression Measuring the Relationships Between the Type of Errors and the Total Completion Time

Variable	B	SE	t	p-value	95% CI
Repetition	-10.812	7.528	-1.44	0.201	-29.233:7.608
Task Addition	12.800	3.924	3.26	0.017	3.199:22.402
Omission	-12.260	5.541	-2.21	0.069	-25.817:1.298
Substitution	103.002	38.708	2.66	0.037	8.288:197.716
Tool Omission	0				
Perseveration	104.090	24.735	4.21	0.006	43.567:164.614
Sequence Failure	-3.049	6.843	-0.45	0.672	-19.794:13.695
Step Recall	26.687	6.923	3.91	0.008	9.991:43.383
Mild Micro Error	59.647	16.849	3.54	0.012	18.419:100.874
Moderate Micro Error	-31.365	16.942	-1.85	0.114	-72.820:10.090
Severe Micro Error	29.200	15.186	1.92	0.103	-7.959:66.359
Efficiency Step	-1.027	9.373	-0.11	0.916	-23.962:21.907
Cleaning Addition	5.270	4.699	1.12	0.305	-6.228:16.769
Organizational Addition	-0.591	3.739	-0.16	0.880	-9.740:8.559

However, as discussed earlier, in the variables measuring the total amount of Task Additions, Sequence Failures, Step Recalls, and Repetitions outliers were found. That is why the previous regression was run again without outliers to see the differences. These appeared to be quite large as in the model without outliers (F(13,2) = 1.05) all variables became nonsignificant, the adjusted R^2 dropped from 0.77 to 0.04 and the model itself also became nonsignificant (p-value from 0.019 to 0.591). Something noteworthy was that the average VIF value measuring multicollinearity was 4.05 for the model, caused by multiple high individual VIF values, like 7.26 for Step Recall and 6.40 for Omission Errors. As the outliers were no measurement error, it was decided to keep the full data set as leading. The model without the outliers can be found in Appendix E.2.

The same model as the one in Table 17 was run again but now with the Total Completion Time Excluding the Duration of the Interruption as the outcome variable ($F(13,6) = 6.51$). This model does not differ greatly from the model displayed in Table 17. They both explained the variance of the outcome variable to a large degree ($R^2_{Adjusted} = 0.77$ for the model in Table 17 and $R^2_{Adjusted} = 0.79$ for the new model) and both have predictive power ($p = 0.019$ for the model in Table 17 and $p = 0.015$ for the new model). All of the individual variables that were significant in Table 17 were also significant in the new model. The model with Total Completion Time Excluding the Duration of the Interruption as the outcome variable can be found in Appendix E.3.

Additionally, it was investigated whether the total MMSE score influenced the resumption lag of participants. Table 18 shows the results for this simple robust regression ($F(1,18) = 1.96$, $p = 0.178$, $R^2 = 0.16$) and indicates that there is no significant relation. The robust option was used as the homoscedasticity assumption was violated. It was also checked whether the removal of the outlier on the Resumption Lag variable made a significant difference and this was not the case.

Table 18

The Results from the Simple Regression with Resumption Lag as Outcome Variable

Variable	B	Robust SE	t	p-value	95% CI
MMSE Score	0.255	0.182	1.40	0.178	-0.127:0.638

Qualitative data analysis

In total 21 interviews were held, of which one with a person with dementia. Thematic analysis resulted in three main themes supported by sub-themes. The themes and sub-themes are listed in Table 19. To support the themes and sub-themes. sometimes quotes from the participants have been used. These are all numbered and the original Dutch versions of the quotes are listed in Appendix F.

Table 19

The Main and Sub-themes of the Interviews held After the NAT

Main Theme	Sub-theme
The interruption in this form did not succeed in impacting the memorization and execution of the NAT.	The interruption was successful in disengaging participants from the NAT. The interruption did not consciously impact the execution of NAT. After the interruption, no difficulties were experienced regarding the continuation of the NAT.
The majority of participants did not use any tactics to remember the instructions. Tactics that were mentioned were the division of the NAT into sub-tasks, holding on to the order of the NAT, or using the objects on the table as cues.	The majority of participants did not use tactics to remember the task. Following the order of the instructions or dividing the NAT into separate sub-tasks were used tactics for remembering the instructions. The items on the table could help in remembering the instructions, but with the current task, they were not truly necessary.
Familiarity and habituation on the task did reduce the resources necessary for memorization and execution of the task, but when habits were incompatible with the task instructions, they could also lead to errors.	Familiarity with the task instructions diminished the need for a tactic to remember them. When instructions were dissimilar from learned habits, more errors were made As the NAT was experienced as fairly simple, no strong emotions were felt.

Main Theme	Sub-theme
	The mixture of target and distractor items caused extra effort as it was not known where the target objects were placed.
The current set-up of this study might limit ecological validity as the lab environment affected participant in various ways different than in their home setting.	By executing the NAT in a controlled lab setting, some situations may arise that are not common in a home setting.
	The execution of the NAT in the lab instead of their home setting impacted the participants and their emotions in various ways.

Theme 1: The interruption in this form did not succeed in impacting the memorization and execution of the NAT.

The goal of the study design was to interrupt participants and see how this interruption impacted their execution and memorization of the NAT. However, after analyzing the results it can be concluded that although the interruption did disconnect the participants from the NAT, it did not succeed in affecting the execution or memorization of the NAT. Also after the interruption, participants did not have difficulties with continuing the NAT.

Subtheme 1.1: The interruption was successful in disengaging participants from the NAT. The goal of the interruption was to let participants stop thinking about the NAT. This was done by asking questions targeting various aspects of memory, like retrospective and prospective memory, and by taking participants away from the table with the objects. After the interviews, it can be concluded that this procedure did work in disconnecting participants from the NAT as all participants indicated not to think about the NAT anymore during the interruption.

Subtheme 1.2: The interruption did not consciously impact the execution of NAT. All 21 participants, including the person with dementia, did indicate that the interruption did not affect the execution of the NAT. Some participants indicated that they thought that this interruption was planned, for example, participant 14 said: *"Yes, I thought like, "Oh is this on purpose?", […] to confuse me so that I forget the task"* (quote 1). Furthermore, participant 20 did indicate not to be influenced by the interruption but when she returned to the table after the interruption, she tried the lid on the thermos bottle. During the interview, this participant indicated that she already saw the lid before the interruption but believed that it did not belong to the thermos bottle. This shows that the interruption still might have a small effect, as it is assumed that without the interruption the lid was not tried anymore. However, this cannot be said with certainty.

Subtheme 1.3: After the interruption, no difficulties were experienced regarding the continuation of the NAT. After answering the question of the interruptions, participants were told that they could move back to the table with all the objects and that they could resume their task. One of the research goals was to see whether participants experienced difficulties with this. During the interview, it was found that participants had no trouble remembering at which point they stopped the task and how they should continue again. They did not use a tactic for this.

Theme 2: The majority of participants did not use any tactics to remember the instructions. Tactics that were mentioned were the division of the NAT into sub-tasks, holding on to the order of the NAT, or using the objects on the table as cues.

Almost all participants indicated that they did not use a tactic to remember the instructions. This might be because they heard the instructions at least twice and were also helped by the visualizations. Some participants did mention memorization techniques, but these were sometimes only mentioned after the experimenter brought them up. The objects on the table were also said to help with remembering how to continue the NAT after the interruption. However, also here participants weakened this claim by saying that this was not truly necessary and they could remember for themselves where to start again. Based on this, it was concluded that, in general, no real strategies were used to remember the instructions of the NAT.

Sub-theme 2.1: The majority of participants did not use tactics to remember the task. Almost all participants indicated that they did not use tactics to remember the instructions of the NAT. This was partly because participants were highly familiar with the components of the task, this was also elaborated on in sub-theme 3.1. Moreover, something to keep in mind is that all participants heard the instructions twice and that in between, they also had to repeat the instructions themselves. Moreover, the visualizations of the task components that were shown also helped with remembering the instructions according to participant 14.

Sub-theme 2.2: Following the order of the instructions or dividing the NAT into separate sub-tasks were used tactics for remembering the instructions. Although most participants indicated to not use any tactics to remember the instructions, some participants did mention some used techniques. One of them was related to the order of instructions. According to multiple participants, keeping the order the same as the instructions made it easier to remember the task. One of those participants was participant 8 who mentioned that she tried to follow the order of the instructions: *"I did try to execute it in the order you said yes"* (quote 2). As reason for doing so, participant 8 explained that it not only helped with remembering the instructions but that the instructions were also something to hold on to, as all the objects on the table were mixed on purpose (quote 3). Another tactic mentioned was to divide the NAT into separate tasks. For example, after pouring in the thermos bottle participant 15 said: "*that is that*" (quote 4), and in the interview it was asked whether that was said because he felt that something was concluded. Participant 15 replied: *"Yes, in fact, yes. Yes that is the first task"* (quote 5). After asking whether the task was seen as separate tasks, participant 15 replied: *"Yes, a little bit yes"* (quote 6). Participants 8 and 9 also mentioned the division of the NAT in sub-tasks. However, this was only mentioned after the experimenter brought it up, which reduced the credibility of their statements.

It can be concluded that some tactics to remember the instructions were mentioned. However, they were not mentioned frequently and they were also sometimes mentioned only after the experimenter brought them up, reducing the strength of this sub-theme.

Sub-theme 2.3: The items on the table could help in remembering the instructions, but with the current task, they were not truly necessary. After finishing the interruption, participants were told that they could go back to the table and that they were allowed to continue the task. In the interview, it was asked whether the objects on the table helped participants remember where they left off. Two participants (8 and 13) mentioned that the objects helped. However, they also indicated that without the objects they would also have known where to start again (quote 7 from participant 8). Participant 15 mentioned that the objects on the table were not necessary for remembering what to do, but that the objects did help (quote 8). Participant 12 also mentioned that seeing the objects on the table helped in remembering where he left off.

The objects on the table could also help in remembering the steps of the tasks, as shown by participants 9 and 15. For example, participant 9 placed the cap on the thermos bottle while doing Task 3 as the cap was not seen earlier. Participant 15 on the other hand, almost forgot to slice the gingerbread and remembered it by seeing it on the table.

Interestingly, the participant with dementia indicated that he did not need the objects on the table to remember where he left off. Moreover, he also indicated that the other objects on the table did not confuse him.

Theme 3: Familiarity and habituation on the task did reduce the resources necessary for memorization and execution of the task, but when habits were incompatible with the task instructions, they could also lead to errors.

The components of the NAT were familiar to almost all participants, only the thermos bottle was less commonly used. On the one hand, this made remembering the instructions easy and the need for a special tactic unnecessary. On the other hand, when the instructions did not match the habits of the participants, participants were prone to make some errors when following their habitual routine. This shows the importance of routine and how difficult it is to stop following it. As the task was such a routine job, it also does not come as a surprise that it did not evoke strong emotions. The only strong emotion felt was surprise that the task was this easy.

Sub-task 3.1: Familiarity with the task instructions diminished the need for a tactic to remember them. When participants were asked whether they had a certain tactic to execute or remember the task, multiple participants indicated that they did not as the instructions were daily actions (participant 12), had a logical order (participant 17), or because the tasks were known (participant 18). Participant 1 indicated that it was easy to remember the instructions as they were in the same order as he always follows, but also because the instructions were not difficult. He said that this was easier than retelling a weird story from a doctor in the geriatric ward (quote 9). Lastly, participant 17 mentioned that it was easier to remember the instructions as the components were familiar to her.

Sub-theme 3.2: When instructions were dissimilar from learned habits, more errors were made. In general, no substantial errors were made while performing the NAT. However, when errors were made, this was sometimes caused by the participants following their habits which interfered with the instructions. For example, participant 15 packed his sandwiches in a sandwich bag before putting it in the lunch box. He explained this behavior by saying that he always does this. Furthermore, participant 7 also packed an extra plastic cup in the backpack, also because she always does so. Another example is participant 13 who forgot to spread butter before spreading the peanut butter. This was also caused by her habit of not putting butter on a sandwich that contains peanut butter. Also, participant 11 mentioned that he automatically started working on the sandwiches, before realizing that first the orange juice needed to be poured in. Participants 7 and 17 both grabbed two bread slices at the same time, which was not instructed. Participant 7 indicated doing so because it was easier as now she only had to open the bread bag once. Participant 17, on the other hand, mentioned again that this was caused by her normal routine. This clearly shows that errors are highly influenced by the habits of participants.

Sub-theme 3.3: As the NAT was experienced as fairly simple, no strong emotions were felt. All participants were asked what their opinion was about the task and how they felt about it, both before and after the task. Most participants answered those questions with "*easy*", "*fun*", "*ordinary*", or "*simple*". Some participants also did not have any emotion at all. After the task, the feelings of "*job well done*" and "*task completed*" were felt regularly. One interesting quote was the one from participant 9 after being asked whether she had a good feeling about completing the task. She mentioned that the task was very easy for her and that she did not feel like she had accomplished something (quote 10). Participant 10 mentioned that he felt "disillusioned" as he questioned himself whether the task was truly this easy (quote 11). It can, therefore, be concluded that no strong emotions were felt, caused by the easiness of the NAT itself.

Theme 4: The current set-up of this study might limit ecological validity as the lab environment affected participant in various ways different than in their home setting.

When doing a study in a lab setting that is unfamiliar to participants, some differences arise compared to doing this study in a familiar home setting. For example, normally at home participants know where to find the objects necessary for task completion. They also know where the objects are that they do not need. As they came to the lab, they did not know what objects were present and where they stood. This might have created some errors that were not made in their home setting. Moreover, in this study, participants had to remember the task instructions. It is assumed that they normally do not follow a strict order in which they make their packed lunch. Coming to the lab, also evoked different emotions in participants, ranging from feeling at home to feeling more stressed. All these factors together limited the ecological validity of this study.

Sub-theme 4.1: The mixture of target and distractor items caused extra effort as it was not known where the target objects were placed. On the table, a mixture of target and distractor objects was placed. This was done to create a natural situation as normally at home, there will also be items that are not necessary for task completion. The difference here was that participants did not know where all the objects were. This resulted in participants who needed to look for the items they needed. For example, participant 10 mentioned that he at first could not find the cheese and asked himself "I needed cheese right?" (quote 12) before finding it. Participant 14 also indicated that she had to look for some objects.

Distractor items were also purposely chosen to distract participants. Participant 10 proved that this worked as he was distracted by the apple juice as its packaging was closer to the thermos bottle than the packaging of the orange juice. Also, participant 14 mentioned that the task contained some issues that could confuse people, for example, the two different bottles (quote 13).

The random placement of all items also caused some participants to reorganize the table, something that has been captured with the Organizational Addition variable. For example, participant 14 mentioned that he had the urge to put objects straight again (quote 14).

Sub-theme 4.2: By executing the NAT in a controlled lab setting, some situations may arise that are not common in a home setting. Participant 3 mentioned that normally at home, a person does not need to follow an exact order. Participant 7 mentioned that all the objects were already there on the table and that she normally needs to grab stuff from the dishwasher. These are two examples of how doing an ADL in the lab differs from doing it at home.

Sub-theme 4.3: The execution of the NAT in the lab instead of their home setting impacted the participants and their emotions in various ways. Being in the lab and performing the NAT evoked various feelings in participants, sometimes also contradicting. For example, participants 6 and 16 mentioned that doing this task made them feel at home, although participant 16 also mentioned that it was a different situation. Still, according to him, the location did not affect the outcome of the task. Participant 4, on the other hand, mentioned that the task was difficult because of the different environment. Participant 17 mentioned something similar as she said that this was not the same as performing the task at home. Being here made this participant a bit shaky and more stressed. Therefore, it can be concluded that performing this ADL in the lab, resulted in different experiences across participants.

Discussion

The goal of this research was to give more insight into how older adults react to an interruption during the performance of an Activity of Daily Living (ADL). This is important as in the near future older adults will need to live longer at home independently and might need assistance with performing ADLs. Assistive technologies can be a solution to this problem if designed properly. As interruptions are a part of daily life, it is important to know how older adults deal with them and whether they can create challenges for older adults. At this moment, scientific knowledge is lacking on the role of interruptions in ALDs, slowing down the potential development of suitable assistive technologies as they can only be well-designed if the problem they solve is clearly defined.

We attempted to contribute to the current scientific understanding of this problem by creating a study that involved an ADL task based on the validated NAT (Schwartz et al., 2002) and combining it with a self-designed interruption. This combination was not done previously with older adults. Moreover, more insight was created by collecting both quantitative and qualitative data and analyzing how they were related, which is something that has not been done in this research area. Finally, it was investigated how older adults with different mental capabilities differed in their execution of the NAT and their reaction to the interruption. In total, 21 participants completed the study, of whom one participant had a diagnosis of dementia. The data of the study was predominantly collected in the Usability lab at the Eindhoven University of Technology.

It can be concluded that although the self-designed interruption did succeed in disengaging participants from the NAT, it had little effect on the planned and actual execution of the NAT. Moreover, differences in cognitive functioning (measured by the MMSE and Working Memory Test), subjective workload (measured by the NASA-TLX), and independence on IALDs (measured by the IADL Questionnaire) could not be used to explain differences in completion times or the number of errors made.

Furthermore, the MMSE score did not relate to the age, gender, or education of participants, their resumption lag, or their answers on the other questionnaires and the Working Memory Test. However, an effect of gender was found on the duration of the task when removing the duration of the interruption. This effect was nonexistent for the duration of the interruption or to the total amount of errors participants made.

Next to this, the subjective workload experienced during the NAT could not be used to significantly predict the duration of the task or the total number of errors made. Although the total duration of the NAT did not relate to the total number of errors made, rejecting a speed-accuracy trade-off, there were some types of errors that significantly impacted the total duration of the task, like making task additions, perseveration errors, substitution errors, step recalls and mild micro errors. It should be noted that there was only one person with a substitution error, but this one error was significantly related to an increase in the total duration of the task of more than 100 seconds. Also, perseveration errors relate to 100 seconds more on the task. Interestingly, of the three types of micro errors, only mild micro errors relate to the duration of the task, whereas moderate and severe micro errors take longer to perform.

Some of the results found seem to contradict earlier research. Starting with the nonsignificant relationship between age, gender, or educational level and the MMSE score, which contradicts previous findings (Fillenbaum et al., 1990; Ishizaki et al., 1998; Rosselli et al., 2006). Scientific literature is, at the same time, not conclusive as well, as Tombaugh et al. (1996) found that men, in general, have lower MSSE scores. Millán-Calenti et al. (2009), on the other hand, found the opposite effect. The latter research also found a relationship between the level of education and the MMSE scores as most of the participants who were cognitively impaired followed education for less than four years (Millán-Calenti et al., 2009). In the current research, almost all of the participants had education for a longer duration, which might explain why the relationship between educational level and MMSE score was absent in this study. The same study by Millán-Calenti et al. (2009) found that cognitive impairment was largest for older people, like people over 85 years old. The current study only included one participant older than 85, which might explain the absence of a relationship between age and MMSE score. The last reason why no relationships have been found might be the narrow distribution of the found MMSE scores, as all scores were 26 or higher, and all participants fell in the healthy category (Zadikoff et al., 2008), which might limit the age, gender, and educational effects.

This last argument might also explain why there was no relationship found between the number of errors participants made and the MMSE score. Weakley and Schmitter-Edgecombe (2019) found that people with MCI took longer to complete the task and made more substitution errors. Healthy older adults, on the other hand, made more omission errors and had a shorter task duration. Based on this, the current research hypothesized that people with MCI would take longer to complete the task after the interruption, would have a longer resumption lag, and would have more (substitution) errors than HOA. As this study, in the end, did not contain people with MCI, it was impossible to test this.

It was, however, analyzed whether the different scores on the MMSE, NASA-TLX, IADL Questionnaires and the Working Memory Test within the HOA group were related to the completion time of the NAT and the number of errors made. As no relationship was found, it might be concluded that differences in task duration and errors made can only be found between groups (MCI and HOA) and not within, as between groups, people significantly differ in their cognitive functioning and independence in executing IADLs. As it is more common to research differences between HOA and people with MCI or dementia, there is currently no research known by the author that analyzed the current research topic within a HOA sample size.

Although not significantly tested, there were differences between the person with dementia and the healthy older adults. On all the tests and questionnaires, the person with dementia scored in the lower segment compared to the HOA. He also took the longest to complete the NAT and made the most errors. These findings are similar to previous findings (Rodríguez-Bailón et al., 2017; Weakley & Schmitter-Edgecombe, 2019), although Weakley and Schmitter-Edgecombe (2019) only tested between HOA and people with MCI. A large difference was found in IADL scores between the person with dementia (0%) and HOA (>75%).

Interestingly, the person with dementia indicated not to be bothered or getting confused by the distractor objects, which seems to contradict our hypothesis and earlier research (Rodríguez-Bailón et al., 2017). It is of course possible that people with dementia are impaired in their ability to reflect on their performance and that their responses do not match their actual behavior. This mismatch between subjective and objective data was found in other literature as well (Most et al., 2012). It is, of course, also plausible that the type of dementia or severity of the disease has an impact on how easily people with dementia get confused by distractor objects. However, as in the current study, only one person with dementia completed the study, the amount of errors made per group (dementia vs HOA) could not be tested and compared with the qualitative answers in the interviews.

Furthermore, this study hypothesized to find a speed-accuracy trade-off in which the task duration of people with higher cognitive functioning, measured by the MMSE, would not be affected by the interruption, compared to a longer task duration of people with a lower

MMSE score. This was based on the paper from Weakley and Schmitter-Edgecombe (2019), which found that an interruption led to longer completion times for people with MCI whereas the completion time of HOA was not affected. Moreover, they found that people with MCI made more substitution errors, whereas HOA made more omission errors. In the current study, no relationship was found between the total number of errors a participant made and the total duration of the task. It is possible that a speed-accuracy trade-off does not exist in a very familiar task. Unlike the study of Weakley and Schmitter-Edgecombe (2019), this research used a very familiar action sequence. Earlier research also showed that when a task was practiced earlier, both speed and accuracy improved (Giovannetti et al., 2007).

Another reason why no speed-accuracy trade-off was found is that the experienced level of rush was, on average, low for this study. It is possible that a speed-accuracy trade-off would emerge if participants were pressured to perform the task as fast as possible. However, when instructing participants to be both as fast as possible and to make as least errors as possible, earlier research also showed no speed-accuracy trade-off (Giovannetti et al., 2007). On the other hand, it can be argued that instructing participants about the speed of the task would contradict the naturalistic aspect of the ADL, as normally people would not be pressured to perform an ADL as quickly as possible. Additionally, it is of course, also possible that people unconsciously increased their speed after the interruption, however, the current study design did not allow to test this, as there was no control condition without interruption present. The qualitative data analysis showed that participants did not consciously alter their strategy or execution of the ADL task after the interruption, which supports the quantitative data.

Although no speed-accuracy trade-off was present, it was found that some errors significantly impacted the total duration of the task. It might be intuitive that performing extra task additions would increase the total duration. Also, recalling steps significantly increased the duration of the task. Based on observations, it is assumed that this is because participants have to think back about what steps they performed and whether they fulfilled all the instructions. Additionally, perseveration errors were also found to increase the total duration of the task. Only two participants made a perseveration error of which one made four instead of two sandwiches. This took a lot of time, which might explain why one perseveration error on average leads to an increase in total task duration of more than 100 seconds.

Most interesting was the fact that each performed mild micro error was related to an increase in task duration of almost one minute, which is substantially longer than the execution of the mild error. Mild micro errors, compared to moderate and severe micro errors, would indicate a good working error managing system as more severe errors were prevented (Divers et al., 2021). It is, therefore, surprising that only mild micro errors relate to the total duration of the task. The study from Divers et al. (2021) was the first known to the author that split micro errors into these three categories. To our knowledge, no other studies have investigated the three different types of micro errors and their effect on task completion time. Earlier research showed that error detection and correction are linked to different cognitive processes, as Bettcher et al. (2008) found a link between executive functioning and error detection and a link between semantic knowledge and error correction. Based on this, it could be argued that people who make mild micro errors have well-functioning executive systems and thus high control over their actions. One part of executive functioning is inhibitory control, which is demonstrated, for example, when a person wants to take more time for a decision and does not immediately react (Diamond, 2013). It is thus plausible that people who make more mild micro errors work slower in general and, therefore, have more time to react to their actions and, therefore, detect when errors are being made.

As already discussed, the current study involved the execution of a highly familiar task (making sandwiches) that participants indicated to execute daily. Research suggests that frequent execution of a task sequence is necessary for the formation of habits (Verplanken, 2006). However, according to Verplanken (2006), frequent behavior is just a precondition for habit formation, and most important factors are the smoothness and the fluency of the habitual behavior. Being smooth and fluent with an action sequence makes it possible for people to execute it without thinking about it and while doing other things at the same time. In the current study, this was also found in the interviews as some participants indicated to have thought about the interruption and whether it was on purpose, showing that they were not thinking about executing the task itself. Wood et al. (2002) also found that habitual behavior was related to less thoughts and attention on the task, but also to less strong emotions. The lack of strong emotions was also found in the current study and supports the notion of familiarity with the tasks indicated by participants.

That participants in the Wood et al. (2002) study thought about something else while executing a familiar task is often interpreted as a sign of automatic behavior (Verplanken, 2006; Wood & Rünger, 2016). Also in this study, some participants described their behavior as automatic. That behavior was automatic might also explain why no tactics were used to remember the instructions and the absence of difficulties when continuing the NAT after the interruption. It might also explain the lack of subjective workload to execute the NAT found in this study. It was hypothesized that people with a higher subjective workload would take longer to complete the task and would make a higher number of errors. However, the collected data did not find such a relationship. It is assumed that this absence can be explained by the simplicity and familiarity of the NAT and the automaticity with which participants executed it. It is expected that when participants execute an unfamiliar task the subjective workload will relate to the performance of the task.

Being familiar with tasks can also have downsides, for example, when deeply ingrained habits are incompatible with the instructions of the task. In those circumstances, habits could lead to action slips, subverting successful task completion (Wood & Neal, 2007). A slip can be defined as the execution of an action that was unintended (Norman, 1981) and is often shown when behavior is automated to such a level that it does not require conscious control anymore (Heckhausen & Beckmann, 1990). In the interview, all participants indicated to be very familiar with the required actions and it was observed that all participants had their own way of doing them. Heckhausen and Beckmann (1990) argue that while doing automated behavior, attention is focused on the environmental aspects that help in maintaining these automated actions.

However, at the same time, this focus on environmental aspects could lead to the rise and activation of other familiar routines and intentions, that might be closely linked to the current performed activity (Heckhausen & Beckmann, 1990). Applied to the current study, this might mean that participants started with executing the instructions, but while doing so, they registered familiar movements and objects, which gave rise to their normal habitual execution of making a packed lunch. This might explain why participants put butter on the gingerbread or forgot to put butter on the sandwich with peanut butter. They saw objects they knew and automatically slipped back into their habits.

These types of errors are called capture errors and can be explained as an error that occurs when the intended schema for action control is overridden by another schema that shares many of the same features with the intended one and is also more frequently used (Norman & Shallice, 1986). The prevention of falling back into a habit requires a significant amount of deliberate attentional resources (Norman & Shallice, 1986). It is possible that because of the simplicity and familiarity of the task, no significant deliberate attention was paid to the task, explaining why some participants fell back into their habits.

Limitations

The largest limitation of the study was the lack of diversity in the group of participants regarding their mental capabilities. It was originally planned to have people with MCI, dementia and healthy older adults. Unfortunately, despite our best efforts, in this short timeframe, it was not possible to arrange this. This made it impossible to test for differences between these three groups. The current study design did also not allow testing for differences within the group of HOA, which can, in hindsight, also be seen as a limitation.

Another limitation is that the sample size was quite small with only 21 participants. Other than the lack of variety in mental capabilities and the small sample, the current sample had a good division of males and females and a relatively large variation in age.

In the current study, the interruption was designed as if the experimenter had forgotten questions that needed to be answered. This made it possible that participants felt bad for the experimenter and out of social desirability, answered that the forgotten questions did not influence how they executed the task. This demand characteristic can be avoided next time by creating another sort of interruption. Options are interruptions that are planned and known for the participant or interruptions that happen outside the power of the experimenter. Moreover, the current study used answering questions as an interruption, it would also be interesting to see whether a task as an interruption has a different effect.

Future studies

As there are only some papers written about how older adults perform ADLs and only one paper involves an interruption, it should not come as a surprise that the current study could not fill all the research gaps. That is why the current study should be considered exploratory. Still, the current study did add knowledge in multiple manners. It, for example, showed how participants reacted to an interruption during the execution of a familiar task and how for HOA no speed-accuracy trade-off was found during the execution of this familiar task. Still, many research gaps remain and future studies are needed to fill those.

Some recommendations for future studies that are related to the current study design are to create a study design that contains three groups (HOA vs MCI vs dementia) and split all three groups into two, creating a between-participant aspect in which half of the participants of each group experience the interruption and half of them do not. This would create the opportunity to research both between and within group effects. Moreover, future research could create a task sequence in which the tasks could be counterbalanced around the interruption to see the exact influence of the interruption on a task. Something not yet studied is the effect of multiple interruptions during an ADL. It would be interesting to see whether a second interruption amplifies the effect of the first one.

Additionally, future research could explore ways to improve the ecological validity of performance-based measures. In the introduction and other research (Giovannetti et al., 2007) it was already elaborated upon on how difficult it is to gather data in a truly ecologically valid (lab-) situation. These difficulties were also experienced in this study. Normally, when a person wants to make a packed lunch, the person knows what objects are available and where they are placed. Moreover, they do not have to follow a certain order in executing the task. This was also indicated by one of the participants.

Moreover, in a normal situation, a person does not have to remember a list of instructions. This also led to problems for one of the persons with dementia who did not finish the study. When explaining the instructions, this person got worried and asked if she had to remember all of the instructions and as she was not capable of doing so, the study had to be stopped before she started the NAT. This suggests that people with dementia might have problems with some task demands that were unrelated to the actual research questions.

Shallice proposes a division between task-resource and task-demand artifacts (Shallice, 1988; Ward, 2015). The first one refers to a situation in which two different tasks use the same cognitive resource, but one task (A) more than the other (B). Damage to this cognitive resource will, therefore, impact task A more than task B. On the other hand, task-demand artifacts explain the difference in task execution by a participant executing one of the tasks sub-optimally. In the current study, participants with dementia may have had problems with the demand of remembering the instructions, but they might not have had problems with making a packed lunch.

Task-demand artifacts could be minimized by clearer instructions (Ward, 2015), which was also experienced during the current study as the person with dementia needed more repetitions of the instructions and needed extra information in general as well. Moreover, one of the participants lost his concentration halfway through the study, which was also a reason to stop the experiment. Based on this, a sufficient concentration span could also be argued to be a task demand. It can be concluded that task-demand artifacts were often a reason for participants with dementia to stop the study. The goal of the study, on the other hand, was to find whether people with different mental capabilities differed in their reaction to an interruption during an ADL task and whether mental capabilities were task resources necessary for successful task completion. Therefore, it can be concluded that the current set-up of performance-based measurements imposed task-demands on participants with dementia, that would not be present in a normal home situation, impacting the ecological validity of the study.

Future research should, therefore, focus on how to research interruptions during ADLs at home. One possibility is to place sensors in participants' homes to create a smart home. The sensors, together with an algorithm, could measure the type and frequency of errors, the duration of an ADL, and also when participants are interrupted. Moreover, participants would have to fill in questionnaires after the performance of an ADL. Post-hoc, this could be combined with subjective data from the participants. This would give insight into when participants face challenges with the execution of an ADL and whether this is related to their measured performance on the ADL. Moreover, it could increase the understanding of what participants experience as an interruption and how this experience affects the execution of an ADL.

To further increase the scope of this future study, it would also be possible to let participants fill in some questionnaires, for example, measuring cognitive functioning or working memory, and test how this relates to the found data. To test this setup it is possible to start with only measuring one ADL, like cooking. The advantage of this study design is that it has much higher ecological validity, as participants are in their own homes and are not imposed with task-demands as described in the current study. Moreover, ADLs can be measured multiple times, instead of only once which is common in lab settings. The downside of the described study design is that it obviously would require a lot of funding.

Implications for assistive technology

One of the goals of this research was to gain insight into what functionalities assistive technologies should possess to help older adults as they need to live longer at home. This research concludes that healthy older adults do not need help to perform their habitual action sequence. However, when participants have to execute actions that are closely related to their habits but slightly differ in requirements, they are more prone to make errors. That is why assistive technology needs to be aware of how older adults normally perform their daily life activities so that when a situation arises that demands alteration of the habitual sequence, this can be highlighted and maybe also repeated throughout the action sequence.

Moreover, the current healthy sample size did not need significant help with memorization of the task instructions. However, based on the three participants with dementia who did not finish the experiment, it seems that this help can become necessary in the future. Therefore, assistive technology needs to be able to help people step by step through the action sequence.

Finally, it was noticed that both persons with dementia who started the NAT looked frequently at the experimenter in the hope of confirmation. That is why assistive technologies should also have functions that signal when a person asks, either verbally or physically, for feedback and functions that provide this feedback. It is assumed that this would be a critical function to help people with dementia as this will make them confident in their daily activities.

Based on the lessons learned, it is suggested to create an assistive technology that consists of sensors placed in peoples' homes combined with a personalized AI algorithm and a tablet screen for communication. The sensors and AI algorithm can be used to learn the habitual execution of ADLs. This habitual execution can then be transformed into a flowchart which can be shown on the tablet. Sensors can also live track ADL execution and communicate this to the tablet so it can show in the flowchart at what step this person is with the execution of the ADL. This way, people can always check at what step they are. The tablet also has a human-like virtual assistant that can communicate and also notices when people look at the tablet for confirmation. As the algorithm knows the habitual execution of ADLs, it can also detect situations when following habits might lead to errors. The virtual assistant can then give prompts to help prevent these errors. The virtual assistant can be personalized, for example, to give prompts at certain moments during the execution of the ADL.

Conclusion

To summarize, it can be concluded that although the self-designed interruption did succeed in disengaging HOA from the ADL task, it did not disrupt the planned and actual execution of the NAT. Healthy participants did not use certain techniques to remember the instructions, likely due to the familiarity with the task. Cognitive functioning, subjective workload, and daily life independence on IADLs did not explain differences in the duration of the task and the number of errors made. Moreover, speed was not related to the number of errors made, rejecting a possible speed-accuracy trade-off. The lack of this trade-off was also explained by the familiarity of the task. In contrast to moderate and severe micro errors, mild micro errors were related to a longer task duration. It is hypothesized that people who make mild micro errors work slower in general and have, therefore, time to realize their mistakes. Overall, our research contributes to our understanding of the role of interruptions when executing ADLs and the extent to which they may disrupt task execution.

References:

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when?: The effects of interruption at different moments within task execution. *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *6*(1), 271–278. https://doi.org/10.1145/985692.985727

Altmann, E., & Trafton, J. (2004). Task interruption: Resumption lag and the role of cues. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*, 43–48. https://escholarship.org/uc/item/18b4r661

Alzheimer Nederland. (n.d.-a). *Alzheimer Café B.O.S.* https://www.alzheimer-nederland.nl/regios/zuidoost-brabant/alzheimer-cafe-bos

Alzheimer Nederland. (n.d.-b). *Alzheimer Café Eindhoven*. https://www.alzheimer-nederland.nl/regios/zuidoost-brabant/alzheimer-cafe-eindhoven

Alzheimer Nederland. (2021). *Factsheet cijfers en feiten over dementie*. https://www.alzheimer-nederland.nl/factsheet-cijfers-en-feiten-over-dementie

Astell, A. J., Bouranis, N., Hoey, J., Lindauer, A., Mihailidis, A., Nugent, C., & Robillard, J. M. (2019). Technology and Dementia: The Future is Now. *Dementia and Geriatric Cognitive Disorders*, *47*(3), 131–139. https://doi.org/https://doi.org/10.1159/000497800

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, *22*(4), 685–708. https://doi.org/10.1016/j.chb.2005.12.009

Bailey, J. (2008). First steps in qualitative data analysis: Transcribing. *Family Practice*, *25*(2), 127–131. https://doi.org/https://doi.org/10.1093/fampra/cmn003

Beaver, J., Wilson, K. B., & Schmitter-Edgecombe, M. (2019). Characterising omission errors in everyday task completion and cognitive correlates in individuals with mild cognitive impairment and dementia. *Neuropsychological Rehabilitation*, *29*(5), 804–820. https://doi.org/10.1080/09602011.2017.1337039

Bemer, A. (2019). Slimmer door de Smartphone? Onderzoek naar Informatieverwerking met Gebruik van Verschillende Devices. *Master Thesis Open Universiteit*.

Berr, C., Wancata, J., & Ritchie, K. (2005). Prevalence of dementia in the elderly in Europe. *European Neuropsychopharmacology*, *15*(4), 463–471. https://doi.org/https://doi.org/10.1016/j.euroneuro.2005.04.003

Bettcher, B. M., Giovannetti, T., MacMullen, L., & Libon, D. J. (2008). Error detection and correction patterns in dementia: A breakdown of error monitoring processes and their neuropsychological correlates. *Journal of the International Neuropsychological Society*, *14*(2), 199–208. https://doi.org/10.1017/S1355617708080193

Bouwstra, H., Smit, E. B., Wattel, E. M., van der Wouden, J. C., Hertogh, C. M. P. M., Terluin, B., & Terwee, C. B. (2019). Measurement Properties of the Barthel Index in Geriatric Rehabilitation. *Journal of the American Medical Directors Association*, *20*(4), 420-425.e1. https://doi.org/10.1016/j.jamda.2018.09.033

Boyle, P. A., Cohen, R. A., Paul, R., Moser, D., & Gordon, N. (2002). Cognitive and motor impairments predict functional declines in patients with vascular dementia. *International Journal of Geriatric Psychiatry*, *17*(2), 164–169. https://doi.org/https://doi.org/10.1002/gps.539

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/https://doi.org/10.1191/1478088706qp063oa

Cahn-Weiner, D. A., Boyle, P. A., & Malloy, P. F. (2002). Tests of executive function predict instrumental activities of daily living in community-dwelling older individuals. *Applied Neuropsychology*, *9*(3), 187–191. https://doi.org/10.1207/S15324826AN0903_8

Cahn-weiner, D. A., Farias, S. T., Julian, L., Harvey, D. J., Kramer, J. H., Reed, B. R., Mungas, D., Wetzel, M., & Chui, H. (2007). Cognitive and neuroimaging predictors of instrumental activities of daily living. *Journal of the International Neuropsychological Society*, *13*(5), 747–757. https://doi.org/https://doi.org/10.1017/S1355617707070853

Caroux, L., Consel, C., Merciol, M., & Sauzéon, H. (2020). Acceptability of notifications delivered to older adults by technology-based assisted living services. *Universal Access in the Information Society*, *19*, 675–683. https://doi.org/10.1007/s10209-019-00665-y

Centraal Bureau voor de Statistiek. (n.d.). *Ouderen*. Retrieved October 4, 2023, from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/ouderen

Centraal Bureau voor de Statistiek. (2020). *Prognose bevolking; geslacht, leeftijd, achtergrond en generatie, 2021-2070*. StatLine. https://opendata.cbs.nl/#/CBS/nl/dataset/84872NED/table?dl=6F90D&ts=1699092943881

Chiu, M.-J., Chen, T.-F., Yip, P.-K., Hua, M.-S., & Tang, L.-Y. (2006). Behavioral and Psychologic Symptoms in Different Types of Dementia. *Journal of the Formosan Medical Association*, *105*(7), 556–562. https://doi.org/https://doi.org/10.1016/S0929-6646(09)60150-9

De Lepeleire, J., Aertgeerts, B., Umbach, I., Pattyn, P., Tamsin, F., Nestor, L., & Krekelbergh, F. (2004). The diagnostic value of IADL evaluation in the detection of dementia in general practice. *Aging and Mental Health*, *8*(1), 52–57. https://doi.org/10.1080/13607860310001613338

Denton, F. T., & Spencer, B. G. (2010). Chronic health conditions: Changing prevalence in an aging population and some implications for the delivery of health care services. *Canadian Journal on Aging*, *29*(1), 11–21. https://doi.org/10.1017/S0714980809990390

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/https://doi.org/10.1146/annurev-psych-113011-143750

Diehl, M., Marsiske, M., Horgas, A. L., Rosenberg, A., Saczynski, J. S., & Willis, S. L. (2005). The Revised Observed Tasks of Daily Living: A performance-based assessment of everyday problem solving in older adults. *Journal of Applied Gerontology*, *24*(3), 211–230. https://doi.org/https://doi.org/10.1177/0733464804273772

Divers, R., Ham, L., Matchanova, A., Hackett, K., Mis, R., Howard, K., Rycroft, S. S., Roll, E., & Giovannetti, T. (2021). When and how did you go wrong ? Characterizing mild functional difficulties in older adults during an everyday task. *Aging, Neuropsychology, and Cognition*, *28*(2), 308–326. https://doi.org/10.1080/13825585.2020.1756210

Edemekong, P., Bomgaars, D., Sukumaran, S., & Levy, S. (2019). Activities of Daily Living. *StatPearls*. https://digitalcollections.dordt.edu/faculty_work/1222

Farias, S. T., Mungas, D., & Jagust, W. (2005). Degree of discrepancy between self and other-reported everyday functioning by cognitive status: Dementia, mild cognitive impairment, and healthy elders. *International Journal of Geriatric Psychiatry*, *20*(9), 827–834. https://doi.org/https://doi.org/10.1002/gps.1367

Fillenbaum, G., Heyman, A., Williams, K., Prosnitz, B., & Burchett, B. (1990). Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly black and white community residents. *Journal of Clinical Epidemiology*, *43*(7), 651–660. https://doi.org/https://doi.org/10.1016/0895-4356(90)90035-N

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Gillham, J. E., Putter, P., & Kash, V. M. (2007). The effects of sad mood on reports of parents’ caregiving behaviors. *Journal of Psychopathology and Behavioral Assessment*, *29*, 125–133. https://doi.org/https://doi.org/10.1007/s10862-006-9031-z

Giovannetti, T., Bettcher, B. M., Brennan, L., Libon, D. J., Burke, M., Duey, K., Nieves, C., & Wambach, D. (2008). Characterization of everyday functioning in mild cognitive impairment: A direct assessment approach. *Dementia and Geriatric Cognitive Disorders*, *25*(4), 359–365. https://doi.org/10.1159/000121005

Giovannetti, T., Bettcher, B. M., Brennan, L., Libon, D. J., Wambach, D., & Seter, C. (2010). Target-related distractors disrupt object selection in everyday action: Evidence from participants with dementia. *Journal of the International Neuropsychological Society*, *16*(3), 484–494. https://doi.org/10.1017/S1355617710000081

Giovannetti, T., Schwartz, M. F., & Buxbaum, L. J. (2007). The Coffee Challenge: A new method for the study of everyday action errors. *Journal of Clinical and Experimental Neuropsychology*, *29*(7), 690–705. https://doi.org/https://doi.org/10.1080/13803390600932286

Giovannetti, T., Yamaguchi, T., Roll, E., Harada, T., Rycroft, S. S., Divers, R., Hulswit, J., Tan, C. C., Matchanova, A., Ham, L., Hackett, K., & Mis, R. (2019). The Virtual Kitchen Challenge: preliminary data from a novel virtual reality test of mild difficulties in everyday functioning. *Aging, Neuropsychology, and Cognition*, *26*(6), 823–841. https://doi.org/10.1080/13825585.2018.1536774

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index):

Results of Empirical and Theoretical Research Sandra. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hartigan, I. (2007). A comparative review of the Katz ADL and the Barthel Index in assessing the activities of daily living of older people. *International Journal of Older People Nursing*, *2*(3), 204–212. https://doi.org/https://doi.org/10.1111/j.1748-3743.2007.00074.x

Heckhausen, H., & Beckmann, J. (1990). Intentional action and action slips. *Psychological Review*, *97*(1), 36–48. https://doi.org/https://psycnet.apa.org/doi/10.1037/0033-295X.97.1.36

Homan, L., Jansen, P., Lubbes, E., & Rijken, E. (2018). Rapportage: Onderzoek naar het huidig gebruik, wachttijden en wachtlijsten Casemanagement Dementie 2018. *Bureau HHM*, 1–39. https://s3-eu-west-1.amazonaws.com/media-assets.hhm.nl/uploads/attachments/LH181386-Rapportage-uitkomsten-onderzoek-huidig-gebruik-wachttijden-en-wachtlijsten-casemanagement-dementie.pdf

Hoonakker, P., Carayon, P., Gurses, A. P., Brown, R., Khunlertkit, A., Mcguire, K., & Walker, J. M. (2011). Measuring workload of ICU nurses with a questionnaire survey : the NASA Task Load Index Measuring workload of ICU nurses with a questionnaire survey : the NASA Task Load Index (TLX). *IIE Transactions on Healthcare Systems Engineering*, *1*(2), 131–143. https://doi.org/10.1080/19488300.2011.609524

Ishizaki, J., Meguro, K., Ambo, H., Shimada, M., Yamaguchi, S., Hayasaka, C., Komatsu, H., Sekita, Y., & Yamadori, A. (1998). A normative, community-based study of mini-mental state in elderly adults: The effect of age and educational level. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *53B*(6), 359–363. https://doi.org/https://doi.org/10.1093/geronb/53B.6.P359

Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., Dubois, B., Eriksdotter, M., Ewers, M., Graessel, E., Kramberger, M. G., Law, E., Mecocci, P., Molinuevo, J. L., Nygård, L., Olde-Rikkert, M. G., Orgogozo, J. M., Pasquier, F., Peres, K., Salmon, E., Sikkes, S. AM., Sobow, T., Spiegel, R., Tsolaki, M., Winblad, B., & Frölich, L. (2015). Mild cognitive impairment and deficits in instrumental activities of daily living: A systematic review. *Alzheimer's Research and Therapy*, *7*(17), 1–20. https://doi.org/10.1186/s13195-015-0099-0

Jones, S., Livner, Å., & Bäckman, L. (2006). Patterns of prospective and retrospective memory impairment in preclinical Alzheimer's disease. *Neuropsychology*, *20*(2), 144–152. https://doi.org/10.1037/0894-4105.20.2.144

Kok, R. M., & Verhey, F. R. J. (2002a). Gestandaardiseerde Mini-Mental State Examination. *Trimbos Instituut*. https://doi.org/https://www.trimbos.nl/docs/b3337fb5-3b14-4832-a4e8-fc7f69e34825.pdf

Kok, R. M., & Verhey, F. R. J. (2002b). *Toelichting scoring gestandaardiseerde MMSE*. 1–2. https://www.trimbos.nl/docs/acb04b78-0a14-4a6a-8961-ac59f7a02980.pdf

Kopelman, M. D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia*, *23*(5), 623–638. https://doi.org/https://doi.org/10.1016/0028-3932(85)90064-8

Kromhout, M., de Klerk, M., & Kornalijnslijper, N. (2018). Veranderde zorg en ondersteuning voor mensen met een beperking. *Sociaal En Cultureel Planbureau*, 1–387. https://www.scp.nl/publicaties/publicaties/2018/06/27/veranderde-zorg-en-ondersteuning-voor-mensen-met-een-beperking

Lai, S. M., Duncan, P. W., Keighley, J., & Johnson, D. (2002). Depressive symptoms and independence in BADL and IADL. *Journal of Rehabilitation Research and*

Development, *39*(5), 589–596.

Lambert, M. A., Bickel, H., Prince, M., Fratiglioni, L., Von Strauss, E., Frydecka, D., Kiejna, A., Georges, J., & Reynish, E. L. (2014). Estimating the burden of early onset dementia; systematic review of disease prevalence. *European Journal of Neurology*, *21*(4), 563–569. https://doi.org/https://doi.org/10.1111/ene.12325

Lawton, M. P., & Brody, E. M. (1969). Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist*, *9*(3_part_1), 179–186. https://doi.org/https://psycnet.apa.org/doi/10.1093/geront/9.3_Part_1.179

Mihailidis, A., Boger, J. N., Craig, T., & Hoey, J. (2008). The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics*, *8*, 1–18. https://doi.org/https://doi.org/10.1186/1471-2318-8-28

Millán-Calenti, J. C., Tubío, J., Pita-Fernández, S., González-Abraldes, I., Lorenzo, T., & Maseda, A. (2009). Prevalence of Cognitive Impairment: Effects of Level of Education, Age, Sex and Associated Factors. *Dementia and Geriatric Cognitive Disorders*, *28*(5), 440–445. https://doi.org/https://doi.org/10.1159/000257086

Ministry of Health Welfare and Sport. (n.d.). *Verpleeghuis*. Retrieved October 16, 2023, from https://www.regelhulp.nl/onderwerpen/ouderen/verpleeghuis

Molloy, D., Alemayehu, E., & Roberts, R. (1991). Reliability of a standardized mini-mental state examination compared with the traditional mini-mental state examination. *American Journal of Psychiatry*, *148*(1), 102–105. https://doi.org/https://doi.org/10.1176/ajp.148.1.102

Molloy, D. W., & Standish, T. I. M. (1997). A guide to the standardized Mini-Mental State Examination. *International Psychogeriatrics*, *9*(1), 87–94.

https://doi.org/https://doi.org/10.1017/S1041610297004754

Monk, C. A., Boehm-Davis, D. A., Mason, G., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors*, *46*(4), 650–663. https://doi.org/10.1518/hfes.46.4.650.56816

Most, E. I. S., Aboudan, S., Scheltens, P., & Van Someren, E. J. W. (2012). Discrepancy between subjective and objective sleep disturbances in early-and moderate-stage alzheimer disease. *American Journal of Geriatric Psychiatry*, *20*(6), 460–467. https://doi.org/https://doi.org/10.1097/JGP.0b013e318252e3ff

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., & Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–1554. https://doi.org/https://doi.org/10.1212/WNL.51.6.1546

Nedjam, Z., Devouche, E., & Barba, G. D. (2004). Confabulation, but not executive dysfunction discriminate AD from frontotemporal dementia. *European Journal of Neurology*, *11*(11), 728–733. https://doi.org/10.1111/j.1468-1331.2004.00981.x

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*(1), 1–15. https://doi.org/10.1037/0033-295X.88.1.1

Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and Automatic Control of Behavior. In *Consciousness and Self-Regulation*, 1–18. Springer. https://doi.org/https://doi.org/10.1007/978-1-4757-0629-1_1

Okken, V., Van Rompay, T., & Pruyn, A. (2012). Exploring space in the consultation room: Environmental influences during patient-physician interaction. *Journal of Health*

Communication, *17*(4), 397–412. https://doi.org/https://doi.org/10.1080/10810730.2011.626498

Patterson, C. (2018). World Alzheimer Report 2018. *Alzheimer's Disease International*, 1–48.

Pérès, K., Helmer, C., Amieva, H., Orgogozo, J. M., Rouch, I., Dartigues, J. F., & Barberger-Gateau, P. (2008). Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: A prospective population-based study. *Journal of the American Geriatrics Society*, *56*(1), 37–44. https://doi.org/10.1111/j.1532-5415.2007.01499.x

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Pinilla, A., Tamayo, R. M., & Neira, J. (2020). How Do Induced Affective States Bias Emotional Contagion to Faces? A Three-Dimensional Model. *Frontiers in Psychology*, *11*, 1–8. https://doi.org/https://doi.org/10.3389/fpsyg.2020.00097

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's and Dementia*, *9*(1), 63-75.e2. https://doi.org/https://doi.org/10.1016/j.jalz.2012.11.007

Rodríguez-Bailón, M., García-Morán, T., Montoro-Membila, N., Ródenas-García, E., Montoro, M. A., & Funes Molina, M. J. (2017). Positive and Negative Consequences of Making Coffee among Breakfast Related Irrelevant Objects: Evidence from MCI, Dementia, and Healthy Ageing. *Journal of the International Neuropsychological Society*, *23*(6), 481–492. https://doi.org/10.1017/S135561771700025X

Rosselli, M., Tappen, R., Williams, C., & Salvatierra, J. (2006). The relation of education and gender on the attention items of the Mini-Mental State Examination in Spanish speaking Hispanic elders. *Archives of Clinical Neuropsychology*, *21*(7), 677–686. https://doi.org/https://doi.org/10.1016%2Fj.acn.2006.08.001

Saito, T., Izawa, K. P., Matsui, N., Arai, K., Ando, M., Morimoto, K., Fujita, N., Takahashi, Y., Kawazoe, M., & Watanabe, S. (2017). Comparison of the measurement properties of the Functional Independence and Difficulty Scale with the Barthel Index in community-dwelling elderly people in Japan. *Aging Clinical and Experimental Research*, *29*, 273–281. https://doi.org/https://doi.org/10.1007/s40520-016-0558-x

Schmitter-Edgecombe, M., Parsey, C., & Cook, D. J. (2011). Cognitive correlates of functional performance in older adults: Comparison of self-report, direct observation, and performance-based measures. *Journal of the International Neuropsychological Society*, *17*(5), 853–864. https://doi.org/https://doi.org/10.1017/S1355617711000865

Schwartz, M. F., Segal, M., Veramonti, T., Ferraro, M., Laurel, J., Ferraro, M., & Buxbaum, L. J. (2002). The Naturalistic Action Test : A standardised assessment for everyday action impairment. *Neuropsychological Rehabilitation*, *12*(4), 311–339. https://doi.org/https://doi.org/10.1080/09602010244000084

Shallice, T. (1988). On Method: Single-Case Studies. In *From Neuropsychology to Mental Structure* (pp. 217–244). Cambridge University Press.

Shigemori, K., Ohgi, S., Okuyama, E., Shimura, T., & Schneider, E. (2010). The factorial structure of the mini mental state examination (MMSE) in Japanese dementia patients. *BMC Geriatrics*, *10*(36), 1–7. https://doi.org/https://doi.org/10.1186/1471-2318-10-36

Shimokihara, S., Tabira, T., Hotta, M., Tanaka, H., Yamaguchi, T., Maruta, M., Han, G., Ikeda, Y., Ishikawa, T., & Ikeda, M. (2022). Differences by cognitive impairment in detailed processes for basic activities of daily living in older adults with dementia.

Psychogeriatrics, *22*(6), 859–868. https://doi.org/10.1111/psyg.12894

Sikkes, S. A. M., De Lange-De Klerk, E. S. M., Pijnenburg, Y. A. L., Gillissen, F., Romkes, R., Knol, D. L., Uitdehaag, B. M. J., & Scheltens, P. (2012). A new informant-based questionnaire for instrumental activities of daily living in dementia. *Alzheimer's and Dementia*, *8*(6), 536–543. https://doi.org/10.1016/j.jalz.2011.08.006

Sikkes, S. A. M., Knol, D. L., Pijnenburg, Y. A. L., De Lange-De Klerk, E. S. M., Uitdehaag, B. M. J., & Scheltens, P. (2017). Supplementary Material for: Validation of the Amsterdam IADL Questionnaire©, a New Tool to Measure Instrumental Activities of Daily Living in Dementia. *Karger Publishers*, 1–4. https://doi.org/https://doi.org/10.6084/m9.figshare.5124436.v1

StataCorp. (2019). *Stata Statistical Software: Release 16.* College Station, TX: StataCorp LLC.

Stopford, C. L., Snowden, J. S., Thompson, J. C., & Neary, D. (2007). Distinct memory profiles in Alzheimer's disease. *Cortex*, *43*(7), 846–857. https://doi.org/https://doi.org/10.1016/S0010-9452(08)70684-1

Stopford, C. L., Thompson, J. C., Neary, D., Richardson, A. M. T., & Snowden, J. S. (2012). Working memory, attention, and executive function in Alzheimer's disease and frontotemporal dementia. *Cortex*, *48*(4), 429–446. https://doi.org/http://dx.doi.org/10.1016/j.cortex.2010.12.002

Sugarindra, M., Suryoputro, M. R., & Permana, A. I. (2017). Mental workload measurement in operator control room using NASA-TLX. *IOP Conference Series: Materials Science and Engineering*, *277*, 1–7. https://doi.org/10.1088/1757-899X/277/1/012022

Thompson, C., Henry, J. D., Rendell, P. G., Withall, A., & Brodaty, H. (2010). Prospective memory function in mild cognitive impairment and early dementia. *Journal of the International Neuropsychological Society*, *16*(2), 318–325. https://doi.org/10.1017/S1355617709991354

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-Mental State Examination (MMSE) and the modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment*, *8*(1), 48–59. https://doi.org/https://psycnet.apa.org/doi/10.1037/1040-3590.8.1.48

Verbeek-Oudijk, D., & Campen, C. van. (2017). Ouderen in verpleeghuizen en verzorgingshuizen. *Sociaal En Cultureel Planbureau*, 1–47. https://doi.org/http://dx.doi.org/10.48592/436

Verbeek-Oudijk, D., & Koper, I. (2021). Het leven in een verpleeghuis. In *Sociaal en Cultureel Planbureau*. https://www.scp.nl/publicaties/publicaties/2021/02/19/het-leven-in-een-verpleeghuis

Verkooijen, L. (2020). *Laatste levensjaren tachtigplussers*. Centraal Bureau Voor de Statistiek. https://www.cbs.nl/nl-nl/longread/statistische-trends/2020/laatste-levensjaren-tachtigplussers?onepage=true

Verplanken, B. (2006). Beyond frequency: Habit as mental construct. *British Journal of Social Psychology*, *45*(3), 639–656. https://doi.org/https://doi.org/10.1348/014466605X49122

Villeponteaux, L., DeCoux, V., & Beardshall, A. (1998). Self-report of functional abilities in older adults with mental retardation: ADLs and IADLs. *Journal of Applied Gerontology*, *17*(1), 53–66. https://doi.org/10.1177/073346489801700104

Waite, L. M., Broe, G. A., Grayson, D. A., & Creasey, H. (2000). Motor function and disability in the dementias. *International Journal of Geriatric Psychiatry*, *15*(10), 897–903. https://doi.org/https://doi.org/10.1002/1099-1166(200010)15:10%3C897::AID-GPS215%3E3.0.CO;2-C

Ward, J. (2015). The Student's Guide to Cognitive Neuroscience. In *The Student's Guide to Cognitive Neuroscience*. Psychology Press.

Watson, P. (2018). *Testing normality including skewness and kurtosis*. MRC Cognition and Brain Sciences Unit. Cambridge University. https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Simon#:~:text=The values for asymmetry and,between -7 to %2B7.

Weakley, A., & Schmitter-Edgecombe, M. (2019). Naturalistic Assessment of Task Interruption in Individuals With Mild Cognitive Impairment. *Neuropsychology*, *33*(1), 1–12. https://doi.org/https://doi.org/10.1037/neu0000481

Wood, W., & Neal, D. T. (2007). A New Look at Habits and the Habit-Goal Interface. *Psychological Review*, *114*(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*(6), 1281–1297. https://doi.org/10.1037/0022-3514.83.6.1281

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314. https://doi.org/https://doi.org/10.1146/annurev-psych-122414-033417

Woods, S. P., Weinborn, M., Velnoweth, A., Rooney, A., & Bucks, R. S. (2012). Memory for intentions is uniquely associated with instrumental activities of daily living in healthy older adults. *Journal of the International Neuropsychological Society*, *18*(1), 134–138. https://doi.org/10.1017/S1355617711001263

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare*, *5*(5), 267–271. https://doi.org/https://doi.org/10.1097/sih.0b013e3181e3f329

Zadikoff, C., Fox, S. H., Tang-Wai, D. F., Thomsen, T., De Bie, R. M. A., Wadia, P., Miyasaki, J., Duff-Canning, S., Lang, A. E., & Marras, C. (2008). A comparison of the mini mental state exam to the Montreal cognitive assessment in identifying cognitive deficits in Parkinson's disease. *Movement Disorders*, *23*(2), 297–299. https://doi.org/https://doi.org/10.1002/mds.21837

Zijlstra, F. R. H., Robert A, R., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, *72*(2), 163–185. https://doi.org/10.1348/096317999166581

Appendix A – Recruitment of Participants

A.1: The Distributed Flyer

Appendix B – The ADL Task

B.1: The Schematic Drawings of the Different Sub-Tasks

Figure B.1.1

The Schematic Drawing of the Steel Thermos Bottle

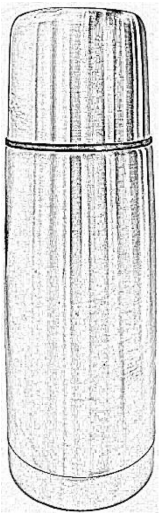


Figure B.1.2

The Schematic Drawing of the two Sandwiches

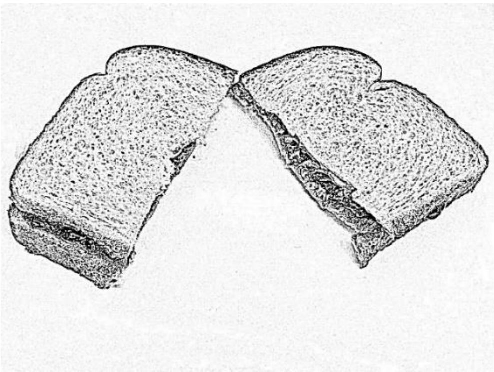
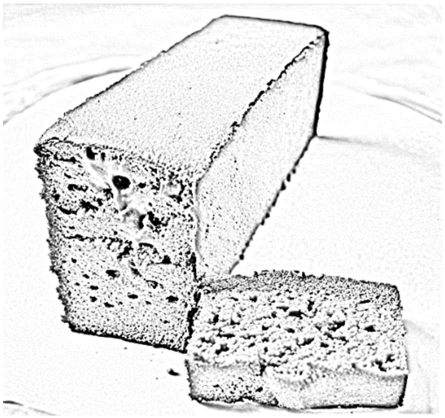


Figure B.1.3

The Schematic Drawing of the Gingerbread



Appendix C – Measurements

C.1: Demographic Questions

- Wat is uw leeftijd? ______________
- Wat is uw geslacht? ______________
- Heeft u kinderen? ______________
    - Hoeveel kinderen heeft u? ______________
- Wat is uw hoogst behaalde onderwijsniveau? ______________
- Wat was uw baan voor u met pensioen ging? ______________
- Bent u een mantelzorger voor iemand? ______________
- Heeft u een officiële vorm van dementie? ______________
    - Welke vorm van dementie heeft u? ______________

C.2: The Items used for the ADL Task

Table C.2.1

All the Objects used in the ADL Task

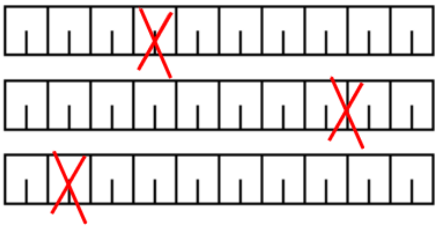
Target object	**Distractor object**
Plate	-
Knife + sharp knife	Spoon, fork, little fork
Orange juice	Apple juice
Steel thermos bottle	Plastic bottle and cup
Peanut butter	Chocolate paste
Cheese	Meats
Butter	-
Bread	Crispbread (Knäckebröd)
Gingerbread	Cookie and chocolate bar
Lunch box	Plastic sandwich bags + extra plastic box

C.3: The Questions asked during the Interruption

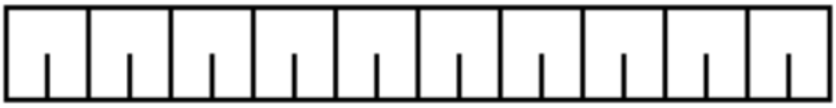
1. Wat heeft u vanochtend ontbeten?
2. Op welk tijdstip eet u normaal gesproken uw lunch?
3. Wat gaat u vanavond als avondeten eten?
4. Hoe zag uw reis naar de universiteit eruit?

C.4: NASA TLX

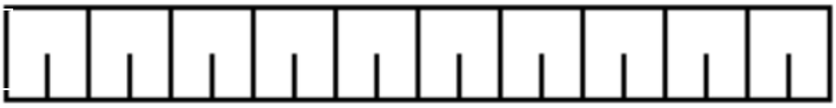
U heeft zojuist de opdracht om een lunchpakket te maken uitgevoerd. Ik wil graag weten hoe u hierop terug kijkt. Daarom staan hier beneden 6 vragen waarvan ik u wil vragen om ze te beantwoorden. Dit doet u door een kruisje te zetten in het voor u meest passende vakje. Hieronder staan 3 voorbeelden van mogelijke antwoorden:



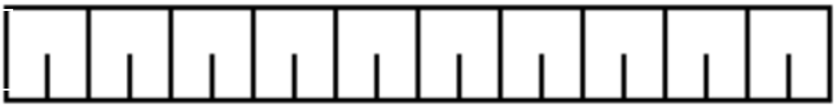
1. Mentaal gezien, hoe veeleisend was de opdracht voor u?

Bijna niet  **Heel erg**

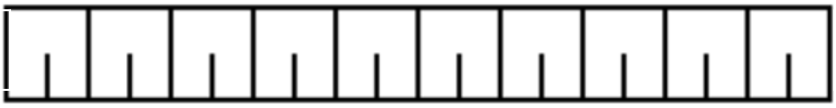
2. Fysiek gezien, hoe veeleisend was de opdracht voor u?

Bijna niet  **Heel erg**

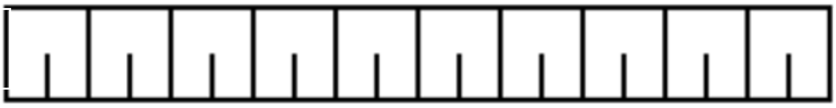
3. Hoe gehaast voelde u zich tijdens het uitvoeren van de opdracht?

Bijna niet  **Heel erg**

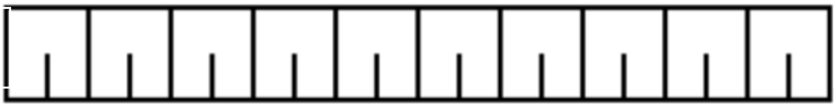
4. Hoe succesvol was u in het uitvoeren van de instructies?

Bijna niet  **Heel erg**

5. Hoe hard moest u werken om deze prestatie te bereiken?

Bijna niet  **Heel erg**

6. Hoe onzeker, ontmoedigd, geïrriteerd, gestrest, en geërgerd was u?

Bijna niet  **Heel erg**

C.5: Semi-structured interview Questions

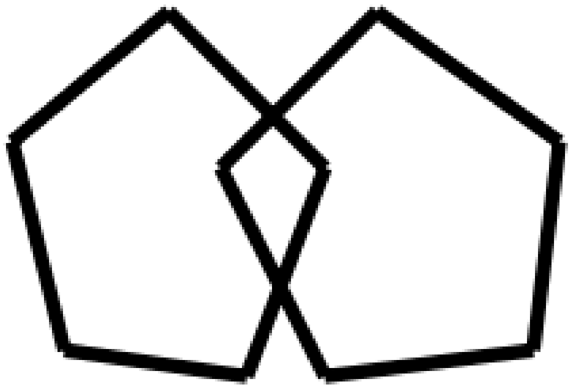
- Wat vond u van de opdracht?
- Wat voor gevoel had u bij de opdracht?
- Zijn de handelingen die u moest doen vertrouwd voor u?
- Had u een bepaalde strategie om de instructies te onthouden?
- Het spijt mij dat ik u moest onderbreken, heeft dat denk u invloed gehad op hoe u de opdracht uitgevoerd heeft?
- Heeft de onderbreking uw strategie om de instructies te onthouden veranderd?
- Was u tijdens de onderbreking nog bezig met de opdracht?
    - Op welke manier?
- Probeerde u tijdens de onderbreking de instructies te onthouden?
    - Hoe probeerde u dit?
- Wat voelde u na de opdracht?
- Wilt u nog iets kwijt over de opdracht of de onderbreking?

C.6: MMSE

		Score
1.	a. Welk jaar is het? b. Welk seizoen is het? c. Welke maand van het jaar is het? d. Wat is de datum vandaag? e. Welke dag van de week is het?	____/5
2.	a. In welk land zijn we nu? b. In welke provincie zijn we nu? c. In welke stad zijn we nu? d. Wat is de naam van dit gebouw? e. Op welke verdieping zijn we nu?	____/5
3.	Ik noem nu drie voorwerpen. Wilt u die herhalen nadat ik ze alle drie gezegd heb? Onthoud ze want ik vraag u over enkele minuten ze opnieuw te noemen. (Noem "appel, sleutel, tafel", neem 1 seconde per woord). (1 punt voor elk goed antwoord, herhaal maximaal 5 keer tot de patiënt de drie woorden weet.	____/3
4.	Wilt u van de 100 zeven aftrekken en van wat overblijft weer zeven aftrekken en zo doorgaan tot ik stop zeg? (Herhaal eventueel 3 maal als de persoon stopt, herhaal dezelfde instructie, geef maximaal 1 minuut de tijd) Noteer hier het antwoord. **of** Wilt u het woord ""worst" achterstevoren spellen?. Noteer hier het antwoord. __________	____/5
5.	Noemt u nogmaals de drie voorwerpen van zojuist. (Eén punt voor elk goed antwoord).	____/3
6.	Wat is dit? En wat is dat? (Wijs een pen en een horloge aan. Eén punt voor elk goed antwoord).	____/2
7.	Wilt u de volgende zin herhalen: " Nu eens dit en dan weer dat " (Eén punt als de complete zin goed is)	____/1
8.	Wilt u deze woorden lezen en dan doen wat er staat'? (papier met daarop in grote letters: "Sluit uw ogen")	____/1
9.	Wilt u dit papiertje pakken met uw rechterhand, het dubbelvouwen en het op uw schoot leggen? (Eén punt voor iedere goede handeling).	____/3
10.	Wilt u voor mij een volledige zin opschrijven op dit stuk papier? (Eén punt wanneer de zin een onderwerp en een gezegde heeft en betekenis heeft).	____/1
11.	Wilt u deze figuur natekenen? (Figuur achterop dit papier. Eén punt als figuur geheel correct is nagetekend. Er moet een vierhoek te zien zijn tussen de twee vijfhoeken)	____/1
	Totale score:	**____/30**

Heeft u deze vragenlijst al wel eens eerder ingevuld? ______

Sluit uw ogen



C.7: Working Memory Test

Table C.7.1

Words Used in the Working Memory Test

Trial	Condition 1 Part A	Condition 1 Part B	Condition 2 Part A	Condition 2 Part B	Condition 3 Part A	Condition 3 Part B
1	Drink Muur Riem	Eend Klei Aan	Noord Hak Wet	Fruit Hond Pas	Hok Geel dag	Eik Bril Heet
2	Gids Klap Snoer	Zuur Plein Maand	Thee Goed Kom	Griek Boer Tong	Vis Bal Huis	Vies Rem Hoog
3	Schaap Recht Lach	Raad Moe Knoop	Pier Gek Hal	Hier Week Pup	Klok Zand Prijs	Peer Laat Klein

Table C.7.2

Numbers used in the working memory test

Trial	Condition 3 – part A	Condition 3 – part B
1	3 64 0 15 30 21 73 14 92 99	12 48 44 9 87 35 76 88 34 3
2	29 50 18 45 8 80 14 44 32	58 94 2 23 86 56 11 30 3 72
3	20 84 13 34 60 14 19 28	15 46 60 76 4 79 44 5 52 1

C.8: IADL Questionnaire

Vaardigheid om telefoon te gebruiken			De was		
	Punten	Antwoord		Punten	Antwoord
Bedient telefoon op eigen initiatief, zoekt en belt nummers op, etc	1		Doet volledig de persoonlijke was	1	
Kan een paar welbekende nummers opbellen	1		Wast kleine items, spoelt kousen etc	1	
Beantwoort de telefoon maar belt zelf niet	1		Alle was moet gedaan worden door anderen	0	
Gebruikt de telefoon helemaal niet	0				

Winkelen			Transportwijze		
	Punten	Antwoord		Punten	Antwoord
Verzorgt zelfstandig alle boodschap behoeften	1		Reist zelfstandig met openbaar vervoer of rijdt met een eigen auto	1	
Winkelt zelfstandig voor kleine aankopen	0		Regelt eigen reizen via taxi, maar gebruikt verder geen openbaar vervoer.	1	
Moet vergezeld worden bij winkeluitjes	0		Reist met het openbaar vervoer wanneer geassisteerd door of vergezeld met een ander.	1	
Volledig niet in staat om te winkelen	0		Reizen gelimiteerd tot taxi of auto met assistentie van iemand anders	0	
			Reist helemaal niet	0	

Eten klaarmaken			Verantwoordelijkheid voor eigen medicatie		
	Punten	Antwoord		Punten	Antwoord
Plant, bereidt en serveert adequate maaltijden zelfstandig	1		Is verantwoordelijk voor het nemen van medicatie met de correcte doseringen op de correcte tijd.	1	
Bereidt adequate maaltijden indien voorzien van ingrediënten.	0		Neemt verantwoordelijkheid wanneer medicatie van te voren is voorbereid in verschillende doseringen	0	
Verwarmt, serveert en bereidt maaltijden voor of bereidt maaltijden voor, maar onderhoudt geen adequaat dieet	0		Is niet in staat in het doseren van eigen medicatie	0	
Moet maaltijden bereidt en geserveerd krijgen	0				

Huishouden			Vaardigheid om met eigen financiën om te gaan		
	Punten	Antwoord		Punten	Antwoord
Onderhoudt het huishouden alleen of met af en toe hulp (bijv, hulp voor zwaar huishoudelijk werk)	1		Beheert financiële zaken zelfstandig (budgetteren, huur betalen, rekeningen betalen, gaat naar de bank), int geld en houdt inkomen bij	1	
Voert lichte dagelijkse taken uit, zoals de afwas en het bed opmaken	1		Beheert dagelijkse aankopen maar heeft hulp nodig met bankieren, grote aankopen etc	1	
Voert lichte dagelijkse taken uit maar kan geen acceptabel hygiëneniveau onderhouden.	1				
Heeft hulp nodig met alle huishoudelijke taken	1		Onbekwaam om geld te beheren.	0	
Neemt aan geen enkele huishoudelijke activiteit deel	0				

Appendix D: Supplementary Descriptive Results Material

D.1: NASA-TLX

Table D.1 shows the results of the individual items of the NASA-TLX. It can be seen that Mental Demand was experienced to be low but did score higher than Physical Demand, which was also expected as the NAT did not require a lot of physical activity. The mean Temporal Demand was 6.5, indicating that participants did not feel rushed during the NAT. The task did cost some effort. Still, the mean Effort score was only 7.38. In general, not a lot of negative emotions were experienced (*M:* 6.13).

It can be seen in Table D.1, that participants in general were quite confident that they fulfilled the instructions of the task, as the mean value is only 9.87. However, this is higher than the means of the other individual items, indicating that they were not convinced that they were perfect in executing the instructions. One value for this item was changed into a missing value as it was suspected that this participant did not realize the reverse scale of the questions. This person indicated on the Performance question that she believed that she failed significantly in executing the instructions. However, in all other questions and the interview, this person indicated to find the task very easy. Also, some participants did not understand the layout of the questionnaire and placed their marks between the answer lines, which is why the minimum value for all items is 2.5.

For all items, the standard deviation is quite similar and not very large. Only for the first item, Mental Demand, the standard deviation is larger (9.24), indicating that participants did vary more in their perception of Mental Demand. This comes back in the maximum value of 45, which is somewhat higher than the maximum value of Physical Demand (35) but much higher than the other maximum values. Only the Performance item has a skewness value within the -2 and 2 range, possibly indicating rejection of normality. The same holds for kurtosis values, as only the Temporal Demand, Performance, and Effort items fall within the

-7 and 7 range. The Shapiro-Wilk and skewness/kurtosis test showed indeed that normality was rejected for all individual items, as for the average. Except for the Performance item, the skewness/kurtosis test did not reject normality (p = 0.16). Removing the outlier did not result in a different outcome.

Table D.1

The NASA-TLX Scores Obtained by Participants

Nasa-TLX	Observations	Mean	SD	Min	Max	Skewness	Kurtosis
Mental Demand	20	7.25	9.24	2.5	45	3.68	15.45
Physical Demand	20	6.25	6.86	2.5	35	3.95	17.12
Temporal Demand	20	6.5	5.03	2.5	22.5	2.16	6.67
Performance	19	9.87	6.74	2.5	25	0.86	2.55
Effort	20	7.38	6.51	2.5	25	2.15	6.10
Frustration	20	6.13	5.10	2.5	25	2.96	10.98
Average	20	7.19	4.23	4.17	23.33	3.00	11.93

Note. For the calculation of the average, all items carried the same weight

The person with dementia had an average score of 19.7, which is higher than the average score of the participants without an official dementia diagnosis, but not the highest score. The reason why the average score is higher than that of the other participants is mostly because the person with dementia experienced a higher Mental Workload (35) and was not confident in his Performance (50). He indicated a small amount of frustration (15), but very low Effort and Physical and Temporal Demand, as these three items all scored 5.

<u>D.2: MMSE</u>

Table D.2 shows all the individual items of the MMSE. There were two items in which most points were lost. First was the Attention/Concentration item, in which participants had to subtract seven from 100 and subtract seven from the resulting number again until five subtractions were made. Participants could earn five points with this item, but the mean score was only 4.45 (*SD: 0.76*, 3-5). Only twelve participants got the full score for this item, three participants only got 3 points, and five participants earned 4 points. The other item that turned out to be difficult was the Delayed Recall item in which participants had to name the three items mentioned right before the Attention Test (*M:* 2.45, *SD:* 0.67). Only 11 participants successfully named all three items, seven participants named two items, and two participants named only one item. On all the other items, participants either received the full score or lost at most one point. On the Naming Task and the Reading a Sentence Task, no participant made an error.

The Skewness/kurtosis test showed that for the total MMSE score normality was not rejected ($p = 0.55$). The same holds for the Attention and Delayed Recall Test, with p-values of 0.13 and 0.18 respectively. The Shapiro-Wilk test gave the same outcome for the total MMSE data ($p = 0.82$). However, it did reject normality for the Attention and Delayed Recall Test. Instead, it failed to reject normality for the Spatial Orientation and the Verbal Comprehension Tests, with p-values 0.17 and 0.17, respectively. Of all 20 participants, only 2 had previous experience with the MMSE, as one participant was present when a loved one filled in the MMSE, and another had filled it in themselves recently. However, as there were only two participants with earlier experience, it could not be tested whether this influenced their MMSE scores.

The participant with dementia had a total score of 25, which is the lowest of all participants. Points were mostly lost at the Delayed Recall Test, as three points were lost at that item. Also, at the Verbal Repetition and Verbal Comprehension items, a point was lost.

Table D.2

The MMSE Scores Obtained by Participants

MMSE	Observations	Mean	SD	Min	Max	Skewness	Kurtosis
Temporal Orientation	20	4.95	0.22	4	5	-4.13	18.05
Spatial Orientation	20	4.9	0.31	4	5	-2.67	8.11
Immediate Memory	20	2.95	0.22	2	3	-4.13	18.05
Attention / Concentration	20	4.45	0.76	3	5	-0.94	2.43
Delayed Recall	20	2.45	0.67	1	3	-0.82	2.53
Naming Task	20	2	0	2	2	.	.
Verbal Repetition	20	0.95	0.22	0	1	-4.13	18.05
Reading a Sentence	20	1	0	1	1	.	.
Verbal Comprehension	20	2.9	0.31	2	3	-2.67	8.11
Writing	20	0.95	0.22	0	1	-4.13	18.05
Constructional Praxis	20	0.95	0.22	0	1	-4.13	18.05
Total	20	28.45	1.05	26	30	-0.42	2.88

Note. The names of the individual items were taken from Shigemori et al. (2010).

D.3: Working Memory Test

Table D.3.1 shows the scores for the three trials individually. In the first and second trials, almost all participants earned the full score of 18, and only one participant in both rounds missed one point. The last trial had a mean score of 13.3 which indicates that this trial was substantially harder for participants.

Table D.3.2 shows the score distribution of the third trial. It shows that three participants remembered 8 of the 18 words and only one participant remembered all of the 18 words. It comes as no surprise that the skewness and kurtosis values for the first two trials show that there was a large and skewed peak. However, the values for the last trial and the total score fall within the earlier predefined range for the skewness en kurtosis values. The Shapiro-Wilk and skewness/kurtosis test indeed showed that normality was not rejected (p >.18).

Table D.3.1

The Scores on the Working Memory Task

Working Memory Test	Observations	Mean	SD	Min	Max	Skewness	Kurtosis
Trial 1	20	17.95	0.22	17	18	-4.13	18.05
Trial 2	20	17.95	0.22	17	18	-4.13	18.05
Trial 3	20	13.3	3.08	8	18	-0.42	2.19
Total	20	49.2	3.27	43	54	-0.57	2.40

Table D.3.2

The Obtained Scores for the Third Trial

Words Remembered	Number of Participants
8	3
10	1
11	1
12	2
13	2
14	4
15	1
16	3
17	2
18	1

D.4: Error Variables

Omission. In total 56 Omission Errors were made, of which 32 took place in Task 1, one in Task 2, five in Task 3, and eighteen after Task 3 (Figure D.4.1). In Task 1, the most common error was forgetting the lid on the thermos bottle, which happened nine times and counted for three Omission Errors on each occasion. The cap on the thermos bottle was only forgotten once, but this participant did include the lid. One participant did not tighten the cap on the thermos bottle, resulting in one Omission Error and another participant broke the orange juice lid, also resulting in one Omission Error as this person failed to properly open the orange juice. One participant forgot to put butter on one sandwich, resulting in the Omission Error during Task 2. Another participant forgot to slice the gingerbread, resulting in the five Omission Errors on Task 3. Failure to pack the thermos bottle and lunch box in the backpack resulted in the omission of six different steps. This happened three times, explaining the eighteen errors after Task 3. The distribution of the Omission Errors is shown in Table D.4.1.

Figure D.4.1

The Number of Omission Errors per Task

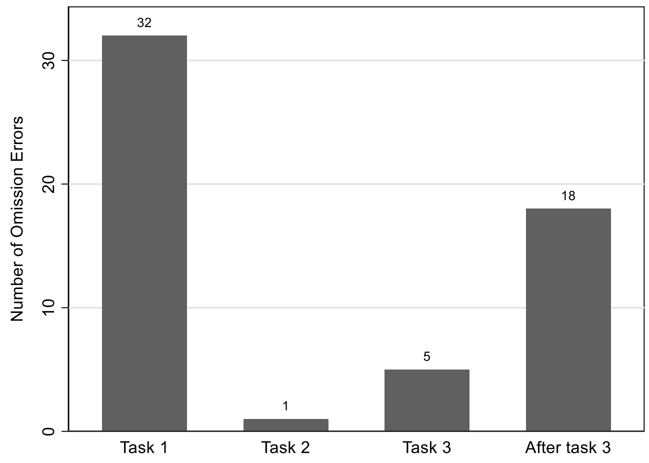


Table D.4.1

The Distribution of Omission Errors

Number of Omission Errors	Frequency
0	5
1	1
3	9
4	1
6	4

Perseveration. In total, three Perseverations were made, two in Task 2, one before and one after the interruption, and one in Task 3. Two of the Perseveration Errors were made by the same participant. This participant made two sandwiches with peanut butter and, therefore repeated the movement of spreading peanut butter on the bread. Furthermore, in Task 3, this participant placed the lid on the lunch box for a second time immediately after doing it for the first time. The other Perseveration Error was made by another participant and involved closing the bread bag twice.

Repetition. In total, 36 Repetitions were made, mostly occurring in Task 2 (Figure D.4.2). This is partly because one specific type of behavior was common among most participants. It was noticed that many participants used the slice of bread to clean the knife so that by sliding it over the bread, the spread on the knife was removed. It was also noticed that some people cleaned their knives and then sliced the slice of bread into two halves and then immediately cleaned their knives again. In this case, cleaning the knife before the slicing was unnecessary as the next movement, slicing the bread, involved the same spread as was used right before the cleaning of the knife and resulted in the same spread on the knife again. That is why it was decided to count one Repetition when participants cleaned their knives, sliced the bread with the same spread, and then immediately cleaned their knives again. This happened 13 times: two participants even did it twice. Another participant was done spreading butter and then decided to spread it some more, resulting in one Repetition. Figure D.4.3 shows that almost half of the participants (8) made one Repetition. Only two participants managed to complete the full task without Repetitions. One participant made four instead of two sandwiches, resulting in six Repetitions: two for grabbing two extra slices, two for buttering both, and two for putting them both in the lunch box (Figure D.4.3). After standardization of the variable, this score came up as a possible outlier. That is why models including this variable were tested twice, once with and once without the outlier.

In Task 1, there was only one Repetition, and this involved closing the thermos bottle again. This Repetition also occurred six times after Task 3, right before placing the thermos bottle into the backpack. One participant even tried to close the thermos bottle twice at this moment. Closing and opening the bread bag again (2 Repetitions), grabbing (1 Repetition), and closing the lunch box (1 Repetition) were the other Repetitions occurring after Task 3. Between Tasks 2 and 3, one participant replaced the sandwich, and another closed the cheese packaging again, resulting in the two Repetitions. In Task 3, two participants replaced the gingerbread slice in the lunch box, resulting in two Repetitions.

Figure D.4.2

The Number of Repetitions per Task

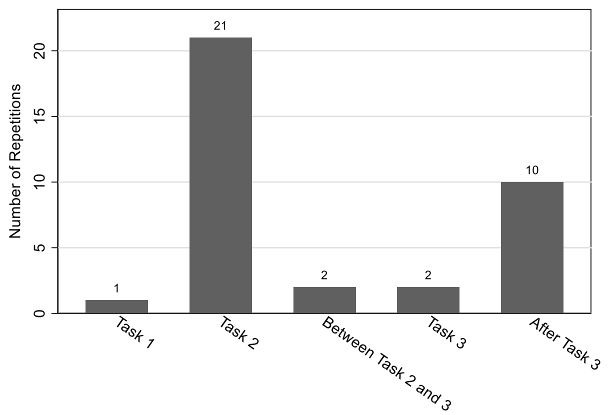
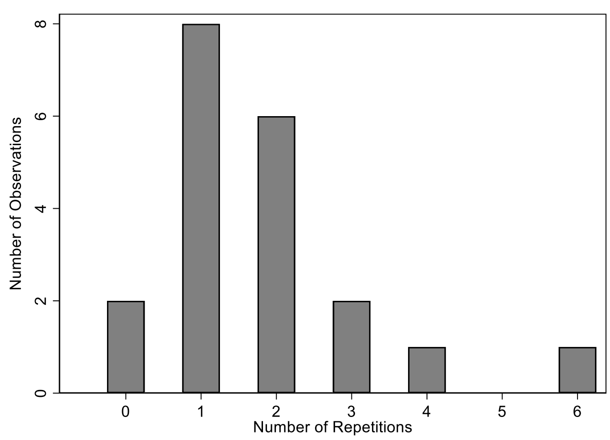


Figure D.4.3

The Distribution of Repetition Errors



Sequence Failures. When arriving at the table for the NAT, one participant started with the second task, mixing the order of the tasks. After performing some steps, it is assumed that this participant realized his mistake as he asked the experimenter whether the order of tasks was important. After the affirmative answer of the experimenter, the participant stopped working on Task 2 and started with Task 1. In the meantime, already six Sequence Failures were made before Task 1 was started, which is shown in Figure D.4.4 and Table D.4.2. Further analysis identified this observation as a possible outlier. That is why models including this variable were tested twice, once with and once without the outlier.

In Task 1, seven Sequence Failures were made, which were all related to the closure of the steel thermos bottle. For example, two participants put the cap already on the bottle and tightened it before pouring the juice, resulting in two Sequence Failures each. Two other participants put the cap on the thermos bottle and another participant put the lid on the empty bottle, all resulting in one Sequence Failure. In Task 2, one person made two Sequence Failures as this participant only grabbed and opened the butter after already making one sandwich, resulting in two Sequence Failures. Another participant placed the cheese slice on the bread before buttering it, which also led to one Sequence Error. Still, the majority of the participants did not make a Sequence Failure.

Figure D.4.4

The Number of Sequence Failures per Task

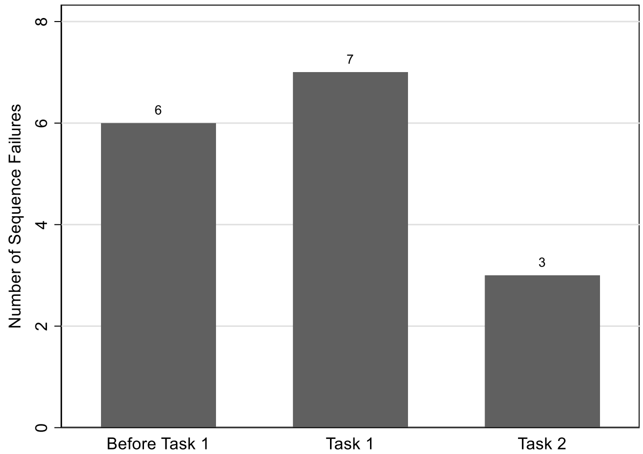


Table D.4.2

The Distribution of Sequence Failures

Number of Sequence Failures	Frequency
0	12
1	4
2	3
6	1

Step Recall. Figure D.4.5 and Table D.4.3 show that one participant made six Step Recalls as he almost forgot to put the steel thermos bottle and lunch box in the backpack. At first, he indicated that he finished the NAT. However, when standing up and getting ready to go to the other table again, he realized that the lunch box and steel thermos bottle were supposed to go into the backpack and he acted accordingly. Standardization showed that this observation could be a possible outlier. Two participants made the same error, as they both realized after finishing Task 3 that the cap present on the table was supposed to be on the steel thermos bottle, resulting in three Step Recalls for both of them. Another participant had already closed the bread bag but later realized a clip was originally used to do so, resulting in that participant closing the bread bag again, but this time with the clip. The majority of participants (16) did not make Step Recalls.

Figure D.4.5

The Number of Step Recalls per Task

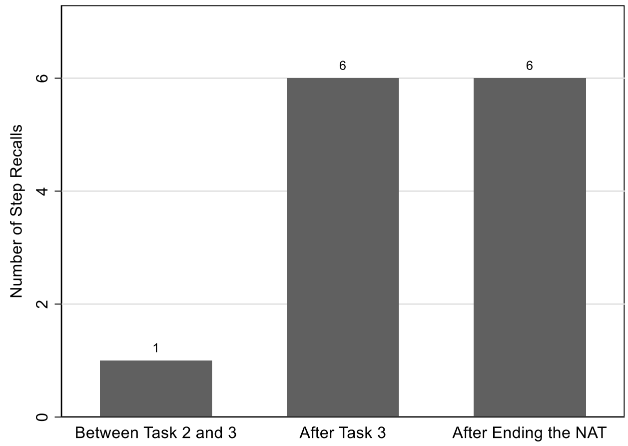


Table D.4.3

The Distribution of Step Recalls

Number of Step Recalls	Frequency
0	16
1	1
3	2
6	1

Task Addition. Figure D.4.6 shows that almost all Task Additions occurred within a task, most of them in Task 2. Most Task Additions that occurred within Task 1 were related to the closure of the thermos bottle before pouring the drink in. By checking whether the lid or the cap fitted the thermos bottle, extra Task Additions were performed. Six participants performed such behavior. In Task 2, some behavior that was reoccurring was to fold the cheese so it fit the bread slice. Four participants showed this behavior. There was one participant who packed the sandwiches in a plastic sandwich bag, which resulted in eight Task Additions. Figure D.4.7 shows that this participant also performed some more Task Additions, resulting in a total of 12 Task Additions. Further analysis showed indeed that this observation was a possible outlier. One participant packed the slice of gingerbread in a sandwich bag, which resulted in six Task Additions. Figure D.4.7 also shows that only two participants completed the task without a single Task Addition. More common was to have three Task Additions, which was done by six participants.

Figure D.4.6

The Total Number of Task Additions per Task

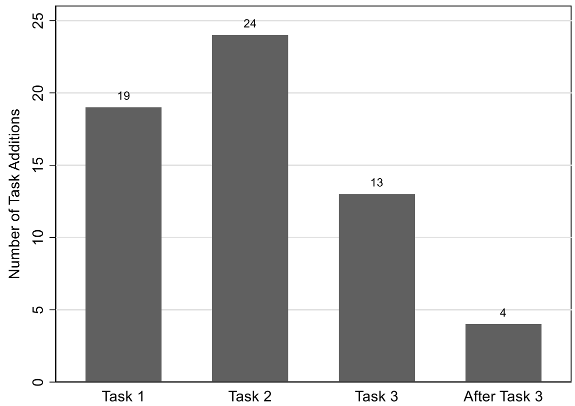
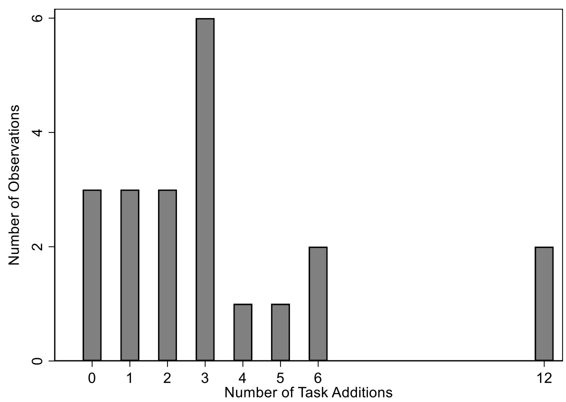


Figure D.4.7

The Distribution of Task Additions

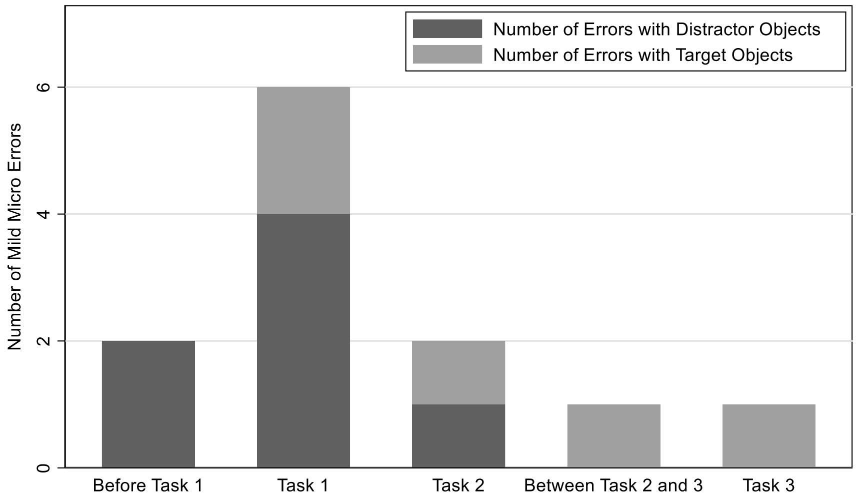


D.5: Micro Errors

Mild Micro Error. In total 12 Mild Micro Errors occurred; these were made by only half of the participants (10). Most of those participants made only one error (8 participants), and only two participants made two Mild Micro Errors. Of the 12 Mild Micro Errors in total, 7 (58%) were made with distractor objects, and 75% occurred within tasks, which is shown in Figure D.5.1.

Figure D.5.1

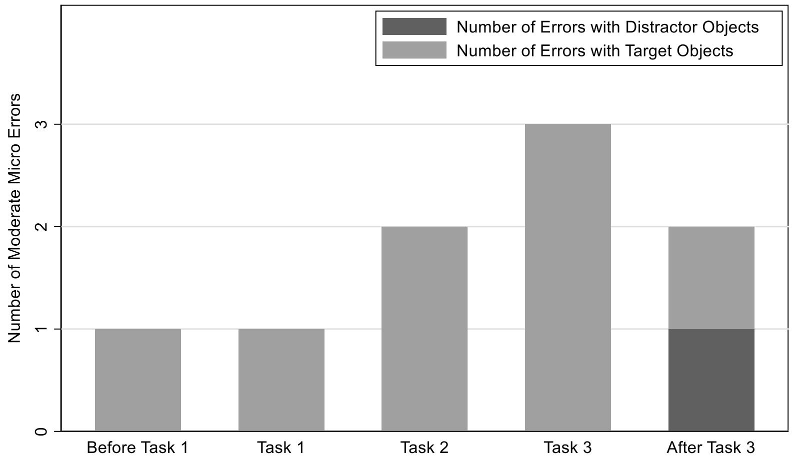
The Number of Mild Micro Errors per Task



Moderate micro error. In total 9 Moderate Micro Errors were made and, as Figure D.5.2 shows, almost all were made with target objects (89%). Only seven participants made a Moderate Micro Error, of which five made one and two participants made two Moderate Micro Errors. Furthermore, most errors occurred within tasks (67%).

Figure D.5.2

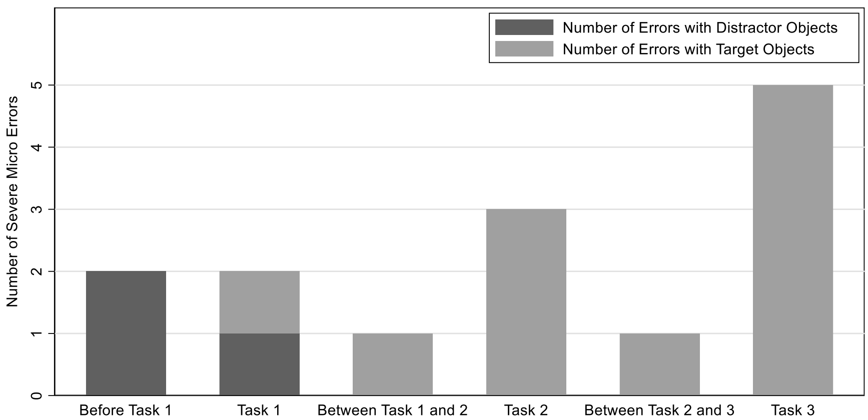
The Number of Moderate Micro Errors per Task



Severe micro error. In total 14 Severe Micro Errors were made of which 11 (79%) with target objects. Moreover, as Figure D.5.3 shows, 10 out of 14 Severe Micro Errors (71%) were made within tasks. More than half of the participants (11) made a Severe Micro Error, as eight participants made one error and three participants made two errors.

Figure D.5.3

The Number of Severe Micro Errors per Task



D.6: Data Enrichment Variables

Efficiency step. Figure D.6.1 shows that most Efficiency Steps occurred in Task 2, which comes from the fact that in this task most opportunities were to combine two steps. In Task 1, the only Efficiency Step that could be made was shaking the orange juice. In Task 2, participants might grab, butter and store two slices at the same time. In Task 3, participants could also perform Efficiency Steps by using the sharp knife to cut the gingerbread. Figure D.6.2 shows that nine participants did not conduct a single Efficiency Step, four participants performed only one, and five participants performed three Efficiency Steps. There is one participant who performed four Efficiency Steps.

Figure D.6.1

The Number of Efficiency Steps Made per Task

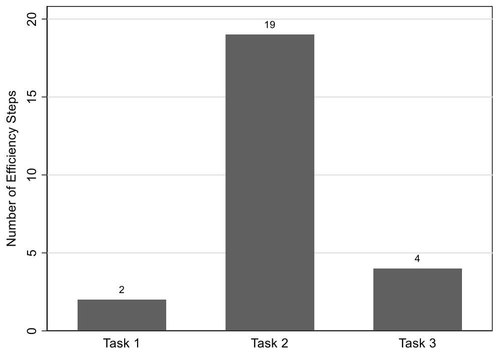
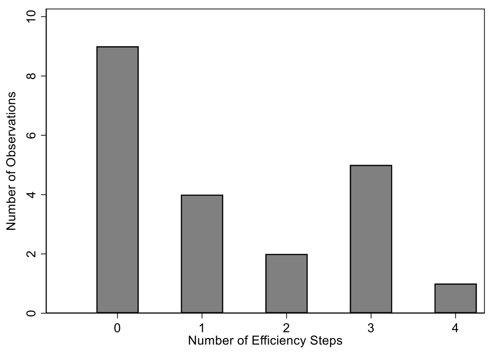


Figure D.6.2

The Distribution of Efficiency Steps



Cleaning addition. As Figure D.6.3 shows, the vast majority of Cleaning Additions were performed in Task 2. This comes as no surprise as in this task, participants open most of the packages, which they later close and they have to make the sandwiches, which leads to the cleaning of the knife on the bread. Figure D.6.4 also shows that participants perform cleaning behavior after Task 2, even after already finishing the task, showing that some participants do care about how they leave the table. Figure D.6.5 shows that all participants did perform some cleaning behavior as the minimum number of performed Cleaning Additions is three. A small minority of the participants (7) performed eight Cleaning Additions, and two participants performed even nine. This shows the large variability in Cleaning Addition amongst participants.

Figure D.6.4

The Number of Cleaning Additions Made per Task

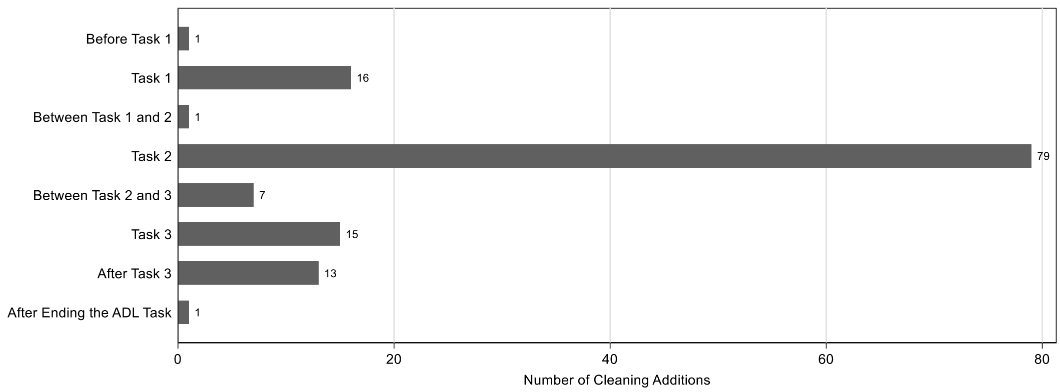
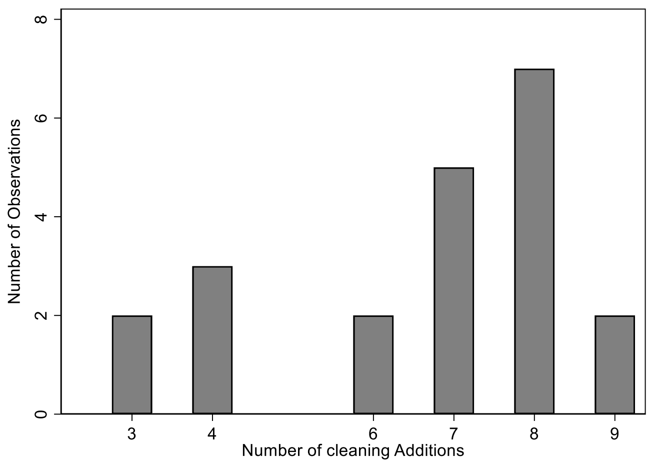


Figure D.6.5

The Distribution of Cleaning Additions



Organizational addition. As Figure D.6.6 shows, most Organizational Additions occurred in Task 2. This could be explained by the fact that in this task objects from the table were most needed, resulting in the need for reorganization of the table. Interestingly, participants also reorganized the table when arriving, before even starting with Task 1 and after completing Task 3. This might indicate that people would like to create a good organization before starting the tasks and after finishing them. Figure D.6.7 shows that there is a large variability among participants in making Organizational Additions as two participants do not make any Organizational Additions and two other participants make eight of such additions. Four participants perform only one Organization Addition, and the other participants vary to a large extent.

Figure D.6.6

The Number of Organizational Additions Made per Task

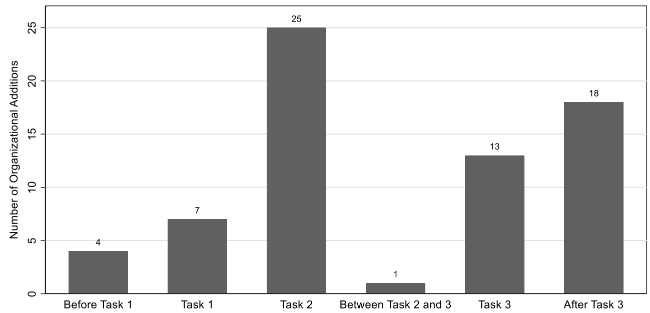
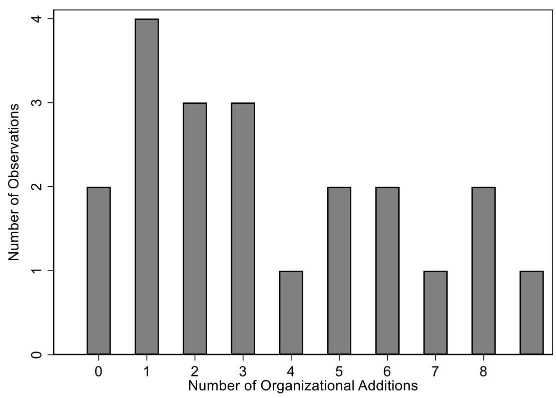


Figure D.6.7

The Distribution of Organizational Additions



Appendix E: Supplementary Hypothesis Testing Material

E.1: Regressions Targeting Before and After Interruption Differences

Table E.1.1 shows the simple regression model that was used to see whether there is a relationship between the score on the MMSE and the Time on Task before the Interruption. When checking assumptions, it was found that both the normality of residuals and homoscedasticity were rejected. By removing the outlier from Time on Task before the Interruption, these assumptions were passed. When looking at Table E.1.1, it can be concluded that there is no relationship between Time on Task before the Interruption and the MMSE score (F(1,17) = 0.00, p = 0.965, $R^2_{adjsuted}$ = -0.06).

Table E.1.1

The Results from the Simple Regression with Time on Task before the Interruption as Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	0.106	2.340	0.05	0.964	-4.831:5.042

Note. One outlier was removed for this model.

Table E.1.2 shows the model that was used to predict whether the total MMSE score had a relationship with Time on Task After the Interruption. When checking assumptions, it appeared that homoscedasticity was rejected, as well as one test that measured the normality of residuals. As there was no outlier in these variables, it was decided to use the robust function when running the model. The outcome of this can be found in Table E.1.2 (F(1,18), p = 0.271, R^2 = 0.07) and shows no relationship between the total MMSE score and Time on Task After the Interruption

Table E.1.2

The Results from the Simple Regression with Time on Task After the Interruption as Outcome Variable

Variable	B	Robust SE	t	p-value	95% CI
MMSE Score	11.971	10.549	1.13	0.271	-10.190:34.133

It was also checked whether the total MMSE score was related to the number of errors made before and after the interruption. It can be concluded that the MMSE score did not have a relationship with the number of errors made either before (Table E.1.3, F(1,18) =0.36, p = 0.558, $R^2_{adjsuted}$ = -0.04) or after the interruption (Table E.1.4, F(1,18) = 1.93, p = 0.182, $R^2_{adjsuted}$ = 0.05). When checking the assumptions for the model displayed in Table E.1.4, it was found that the normality of the residuals was rejected. Further analysis showed that the outcome variable measuring the total number of errors after the interruption was not normally distributed. This was solved by a square root transformation.

Table E.1.3

The Results from the Simple Regression with the Total Number of Errors Before the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	0.317	5.31	0.60	0.558	-0.799:1.454

Table E.1.4

The Results from the Simple Regression with the square root transformation of Total Number of Errors After the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
MMSE Score	0.276	0.198	1.39	0.182	-0.141:0.692

Finally, it was also analyzed whether a speed-accuracy trade-off was present before or after the interruption. When analyzing the speed-accuracy trade-off before the interruption, the normality of the residuals assumption was rejected. Removing the outlier on the Time on Task before the Interruption solved this. Figure E.1.5 shows the speed-accuracy trade-off before the interruption ($F(1,17) = 0.09$, $p = 0.768$, $R^2_{adjsuted} = -0.05$). As can be seen, no speed-accuracy trade-off was present.

Table E.1.5

The Results from the Simple Regression with Time on Task Before the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
Total Number of Errors Before the Interruption	0.3	1.002	0.30	0.768	-1.813:2.413

Figure E.1.6 shows the speed-accuracy trade-off after the interruption. As homoscedasticity was rejected, a robust regression was performed. Figure E.1.6 shows that no speed-accuracy trade-off was present after the interruption ($F(1,18) = 1.55$, $p = 0.229$, $R^2 = 0.13$).

Table E.1.6

The Results from the Simple Regression with Time on Task After the Interruption as the Outcome Variable

Variable	B	Robust SE	t	p-value	95% CI
Total Number of Errors After the Interruption	3.329	2.670	1.25	0.229	-2.282:8.939

E.2.: The Regression Model Testing the Relationship between Total Completion Time and Type of Errors Without Outliers

Table E.2.1

The Results of the Regression without Outliers and the Total Completion Time as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
Repetition	-16.446	19.476	-0.84	0.487	-100.246:67.353
Task Addition	8.489	14.234	0.60	0.611	-52.755:69.733
Omission	-8.519	15.204	-0.56	0.632	-73.937:56.899
Substitution	90.932	78.264	1.16	0.365	-245.809:427.673
Tool Omission	0				
Perseveration	75.826	86.706	0.87	0.474	-297.241:448.893
Sequence Failure	0.653	26.416	0.02	0.983	-113.007:114.312
Step Recall	21.092	34.277	0.62	0.601	-126.389:168.574
Mild Micro Error	57.842	41.830	1.38	0.301	-122.136:237.820
Moderate Micro Error	-27.170	42.225	-0.64	0.586	-208.852:154.511
Severe Micro Error	20.665	22.327	0.60	0.607	-126.644:167.973
Efficiency Step	0.259	22.327	0.01	0.992	-95.808:96.327
Cleaning Addition	8.152	11.302	0.72	0.546	-40.479:56.783
Organizational Addition	-2.540	9.788	-0.26	0.820	-44.656:39.577

E.3: Regression Analysis Testing the Relationships Between Type of Errors and Total Completion Time Excluding the Duration of the Interruption

Table E.3.1

The Results from the Regression with Total Completion Time Excluding the Duration of the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
Repetition	-9.973	6.035	-1.65	0.149	-24.740:4.793
Task Addition	10.777	3.146	3.43	0.014	3.080:18.474
Omission	-5.990	4.442	-1.35	0.226	-16.858:4.878
Substitution	68.777	31.030	2.22	0.069	-7.150:144.704
Tool Omission	0				
Perseveration	69.768	19.828	3.52	0.013	21.250:118.286
Sequence Failure	-1.804	5.486	-0.33	0.753	-15.227:11.619
Step Recall	33.934	5.470	6.20	0.001	20.549:47.318
Mild Micro Error	56.531	13.507	4.19	0.006	23.481:89.582
Moderate Micro Error	-15.396	13.581	-1.13	0.300	-48.629:17.836
Severe Micro Error	6.915	12.174	0.57	0.591	-22.873:36.702
Efficiency Step	17.229	7.514	2.29	0.062	-1.156:35.614
Cleaning Addition	-2.026	3.767	-0.54	0.610	-11.243:7.192
Organizational Addition	2.049	2.997	0.68	0.520	-5.286:9.383

The model in Table E.3.1 was also run without all the outliers. This output can be seen in Figure E.3.2. Removal gave a similar situation as previously seen, as all individual variables became nonsignificant and the model itself also lost its predictive power (F(13,2) =

4.60, p = 0.192). On the other hand, the adjusted R^2 remained high (0.76). However, also for this model, the average VIF value was high (4.05), just like some individual VIF values, like 7.26 for Step Recall and 6.40 for Omission Errors. That is why, the model with the complete data set is again leading.

Table E.3.2

The Results from the Regression without Outliers and Total Completion Time Excluding the Duration of the Interruption as the Outcome Variable

Variable	B	SE	t	p-value	95% CI
Repetition	-1.681	6.130	-0.27	0.810	-28.056:24.693
Task Addition	8.132	4.480	1.82	0.211	-11.144:27.407
Omission	-0.353	4.785	-0.07	0.948	-20.942:20.237
Substitution	68.036	24.632	2.76	0.110	-37.947:174:018
Tool Omission	0				
Perseveration	95.086	27.289	3.48	0.073	-22.330:212.502
Sequence Failure	-3.070	8.314	-0.37	0.747	-38.842:32.703
Step Recall	15.047	10.788	1.39	0.298	-31.368:61.466
Mild Micro Error	51.112	13.165	3.88	0.060	-5.533:107.756
Moderate Micro Error	5.780	13.290	0.43	0.706	-51.400:62.961
Severe Micro Error	-5.743	10.775	-0.53	0.647	-52.106:40.619
Efficiency Step	9.368	7.027	1.33	0.314	-20.868:39.603
Cleaning Addition	-0.554	3.557	-0.16	0.891	-15.859:14.752
Organizational Addition	5.859	3.081	1.90	0.198	-7.396:19.114

Appendix F: Quotes from the Interview

Table F.1

Quotes from the Interview

Quote	Subtheme	Original Dutch Quote	Translated quote
1	1.2	*"Ja, ik dacht van, "oh is dit bewust?. Dacht ik even van uh. Om mij in verwarring te brengen zodat ik dan de opdracht zou vergeten".*	*"Yes, I thought like, "Oh is this on purpose?", […] to confuse me so that I forget the task".*
2	2.2	*"Ik heb wel geprobeerd om ze in volgorde zo als je ze gezegd [E: Ja] hebt, om [E: Ja] ze zo uit te voeren".*	*"I did try to execute it in the order you said yes".*
3	2.2	*"Ook om te onthouden, maar [E: Ja] ook om zeker op zo'n tafel waar je echt alles bewust uh, door elkaar hebt liggen [E: Ja] is de opdracht natuurlijk wel een houvast*)".	Explanation done in-text.
4	2.2	*"Dat is dat".*	*"That is that".*
5	2.2	*"Ja, in feite wel. [E: Ja]. Ja. [E: Dus u ziet het als (?)] ja. Ja, dat is de eerste taak, dus uh. [E: Ja] Ja".*	*"Yes, in fact, yes. Yes that is the first task"*
6	2.2	*"Uhhhh Jaaa (1.5) Beetje wel ja"*	*"Yes, a little bit yes"*
7	2.3	*"Nou ja, die boterham lag er natuurlijk dus [E: Ja] dat was duidelijk. Maar ik denk [E: Ja] dat ik het anders ook nog wel had geweten hoor".*	Explanation done in-text.
8	2.3	*"Hhhmm. Nee. Nou ja, het helpt wel, natuurlijk he [E: Ja]. Ja. Ja)".*	Explanation done in-text.
9	3.1	*"Dat is wat anders als een vreemd verhaaltje horen bij de Geriater en dat na [E: ja] proberen te vertellen".*	Explanation done in-text.
10	3.3	*"Ja, maar uh, het is uh, voor mij heel makkelijk dit. 'T is niet alsof ik een prestatie heb zitten leveren".*	Explanation done in-text.

Quote	Subtheme	Original Dutch Quote	Translated quote
11	3.3	*"Uuhmm, (..) Uhh, ja ik mag het woord ontgoocheling niet gebruiken, maar ik denk (beiden lachen) ik denk: "is dit nou echt waar, wat is, is het, is het zo simpel?".*	Explanation done in-text.
12	4.1	*"Alleen, ik, ik, moest even kijken uhhh, (.) uh, waar ligt de kaas want die z, eh had ik (?) even over het hoofd gezien. [E: hm] Toen dacht ik wel, "ik moest toch kaas?". Maar toen in één keer denk ik "ach (?), hier ligt ie", dus*".	[…] *I needed cheese right?"*
13	4.1	"*Zaten wel dingen waar je denkt, uhhh, daar zou je mensen uhh, in verwarring kunnen brengen, omdat er twee verschillende flessen stonden en zo. Maar het was wel duidelijk een thermos kan. [E: Ja]. Das toch wel net iets anders dan een uh water fles".*	Explanation done in-text.
14	4.1	*"Nee, waar ik wel bewust van ben, dat ik daar dan, heb ik wel de nijging uh, dat ik alles zo weer recht te zetten. Ja, dat is een beetje.. (1.7) [E: Ja]. Dat is gewoon een he, eigen"*	Explanation done in-text.