Working Paper 2011-3B

MIT AgeLab Delayed Digit Recall Task (n-back)

By Bruce Mehler, Bryan Reimer & Jeffery A. Dusek

Original Release: May 11, 2011
Update B: June 28, 2011

Abstract: This document describes both subject training and the experimental administration of the auditory presentation – verbal response delayed digit recall task (n-back) used by the MIT AgeLab in a series of simulation and on-road driving studies. The full stimulus item set, training materials and instructions are provided to assist other researchers who are interested in using the task and methodology in other work.

Introduction

This document describes both subject training and the experimental administration of the auditory presentation – verbal response delayed digit recall task (n-back) used by the MIT AgeLab in a series of simulation and on-road driving studies. The same item content has been used consistently starting in the AgeLab simulator in 2006 (Mehler, Reimer, Coughlin & Dusek, 2009), a pilot on-road study in 2007 (Reimer, 2009; Reimer, Mehler, Coughlin, Godfrey & Tan, 2009; Mehler, Reimer & Wang, 2011), methodological studies in the simulator (Wang, Reimer, Mehler, Zhang, Mehler & Coughlin, 2010), a large on-road study in 2008 (Mehler, Reimer & Coughlin, 2010; Reimer, Mehler, Wang & Coughlin, 2010) as well as subsequent projects that have not yet been reported in the literature. In addition to studies conducted at the AgeLab, a study using the protocol described here has been carried out by colleagues in Korea (Son, Mehler, Lee, Park, Coughlin & Reimer, 2011). The full stimulus item set, training materials and instructions are provided to assist other researchers who are interested in using the task and methodology in other work. In addition, background on the conceptualization and development of the task is presented.

Background on the Form of N–back Employed & Task Difficulty Levels

The form of the n-back task used in these experiments may be best understood by referring directly to the instructions that were used to present the task to subjects during the training period. These are reproduced in Appendix A. As discussed in more detail shortly, these tasks differ somewhat from "n-back" matching tasks that can also be found in the literature.

The delayed response task (n-back) used in the aforementioned AgeLab studies consists of simple auditory stimuli that the driver listens to and repeats back following specific rules. The auditory attention and memory components of the task draw on many of the same cognitive resources utilized when engaging in an externally paced task such as responding to a cell phone call or interacting with an in-vehicle device that uses auditory prompts or control commands. Similarly, it draws on cognitive resources that are utilized for less structured interactions such as attending to and maintaining a conversation with a passenger. The structure of the task allows the total mental workload to be systematically varied across a very mild task demand (0-back) through a moderate level (1-back) and a high level of task demand (2-back).

At the lowest workload level (0-back), participants were required to respond to each of the randomly ordered auditory stimuli (single digits 0–9) by immediately repeating out loud the last number presented. As detailed in Appendix A, the task is explained to participants as follows:

> *"During this task, I will read a list of ten single digit numbers. As I read each number, you are to repeat out loud the last number that you've heard. For example, if I were to say the number 3, you would say 3; then if I said 2, you would say 2; then if I said 6, you would say 6, and so on."*

While the 0-back appears to be a minimally demanding task, we believe that inclusion of a seemingly very low demand level is critically important in work considering scaled demand. This is particularly true in work involving secondary tasks where the addition of relatively modest demands can result in easily measureable effects. In both the simulation and on-road driving studies, statistically significant increases in physiological arousal were obtained when participants engaged in the 0-back task (Mehler et al. 2009; 2010; 2011, Reimer et al. 2009; Son et al., 2011). Similarly, marked changes in visual scanning behavior can be observed, particularly under actual driving conditions (Reimer 2009; Reimer et al., 2010).

At the moderate level (1-back), participants were required to respond with the next-to-last stimulus that was presented:

> *"The second version of the task is called the 1–back, which simply means that as I read each list of ten numbers, you are to repeat out loud the number before the last number that you heard. For example, if I said 3, you would say nothing, then if I said 2, you would say 3, then if I said 6, you would say 2, and so on."*

The 1-back task clearly adds to the basic demand of the easier 0-back. Following Wickens' (2002) description of task stages, both tasks involve the same sequence of a sensory processing stage along the auditory dimension, investment of resources in the perception of the auditory content, holding the perceived content in working memory, and investment of resources in selection of a verbal response mode and execution of that response. The 1-back adds to the demand of taking an item into working memory by requiring that the earlier item continue to be maintained long enough to be processed and executed on as the appropriate response.

In the most difficult level (2-back), participants responded with the second-to-last stimulus:

> *"The final version of the task is called the 2–back, which simply means that as I read each list of ten numbers, you are to repeat out loud the number that was read two numbers ago. For example, if I were to say the number 3, you would say nothing, then if I said the number 2, you would say nothing, then if I said 6, you would say 3, then if I said 7, you would say 2, and so on."*

The 2-back task adds to the overall demand not only by adding a third item that must be maintained in working memory but also increments modestly but meaningfully the task of maintaining the correct sequencing of the three items while the response is processed and executed.

Zeiltin (1993; 1995) demonstrated the utility of the 1-back form of the task under actual driving conditions and our group has used the 0-, 1-, and 2-back forms under simulation and on-road conditions as noted previously. In his 1993 paper, Zeiltin lists a number of requirements and features of an ideal subsidiary task for studying workload (such as interacting minimally with the primary task, require minimal learning, require minimal equipment, be easy to score) and argued that delayed digit recall task is a good candidate for meeting the majority of these criteria after considering a range of tasks that might administered in the context of driving research.

Having mentioned Zeiltin's work, it is worth keeping in mind the differing ways in which secondary tasks are typically employed. Zeiltin highlights an approach that uses performance on the secondary task as an indirect measure of workload. If the demand associated with a primary task increases, it should eventually impact performance on the secondary task. This model assumes that individuals have a finite amount of resources that can be invested in overall task performance and that as demand increases, primacy will be given to the primary task and this will result in performance degradation in the secondary task. There are some limitations to using secondary task performance as a workload measure. For example, it has been predicted under the multiple resource model of information processing that there should be little or no initial impact on secondary task performance if the primary and secondary tasks involve different sensory processing and response channels (Jamson & Merat, 2005; Wickens, 1984; Wickens & Liu, 1988). Nonetheless, monitoring changes in secondary task performance can provide a useful methodology if care is taken in selecting demand characteristics and a demand level that is appropriate for a particular research question.

In our work, the n-back task has been used to induce varying levels of demand so that the impact on participants can be observed. In this application, the scalability of the task is one of its most attractive features.

In selecting secondary tasks and in interpreting results, it is important to consider not only the objective difficulty of the task but also the nature of the resources required to carry out the task. Tasks can vary significantly in the extent to which they place demands on different mental resources, e.g. perceptual processing, short term memory, visual spatial manipulation, etc. The form of the delayed digit recall task presented here is particularly attractive since the auditory presentation – verbal response format does not directly interfere with the visual-manipulative demands of the primary driving task.

Because the difficulty of the task is defined by how many numbers back in the presentation sequence must be kept in working memory, the task can be classified as an “n-back” task. It is useful to note that this form differs from the n-back task frequently used in neuropsychological research. The latter form typically requires participants to indicate whether a currently presented stimulus is the same as a target stimulus presented n-trials previously (Owen, McMillan, Laird, & Bullmore, 2005); this is a more difficult task for a given level of “n” since it involves holding items in working memory, making target matching decisions and, in some versions, shifting targets as the task proceeds. The 0-back and 3-back tasks used in Lenneman et al.’s driving simulation study (2009) were of the target matching form and involved single letters presented visually as overhead signs. These distinctions are important in considering various aspects of demand created by a task (i.e. auditory vs. visual presentation, recall vs. recall and matching); nonetheless, the basic principle that task demand increases with the “n” level applies across studies.

Early Protocol Using Fixed Order Demand Levels

In the initial phase of this work that was carried out in the AgeLab simulator beginning in 2006, the three levels of the task were presented in a fixed order of difficulty starting with the low demand level (0-back), progressing to the medium demand level (1-back) and concluding with the high demand level (2-back). This was done intentionally to observe participants’ reactions to a continually building level of stress coming from both the increasing degree of objective demand and from sustained effort; no recovery periods were provided between tasks. In addition, no pre-experimental training in the tasks was provided. Training instructions and practice sets were introduced while the subject was actively driving the simulator and had accumulated 18 minutes of total simulation driving experience. Details of the protocol are provided in Mehler et al. (2009). This same basic protocol was extended to an actual on-road driving experiment in 2007 (Reimer, 2009; Reimer, et al., 2009).

A primary goal of the early simulation study was to identify minimally invasive physiological measures that could be practically employed to detect increasing stress levels in participants that were actively driving the simulator (as opposed to sitting quietly in a standard laboratory setting). The protocol worked well for that purpose; however, it also left open questions related to order effects in interpreting the relative change in physiological measures between demand levels. This, and other factors, led to the development of a revised protocol that pre-training subjects in the secondary task prior to assessing performance while driving, presenting the demand levels in random order across subjects to control for and assess order effects, and introduction of 2 minute long recovery intervals between the different demand levels. This revised protocol is documented in detail in the remainder of this paper.

Protocol Based on Pre–Training and Random Order Presentation of Demand Levels

The specific protocol presented here was used in the 2008 on-road study (Mehler, Reimer & Coughlin, 2010; Reimer, Mehler, Wang & Coughlin, 2010). Replication or other research building on this work can use either or both of these papers as appropriate citations. This protocol was recently employed by a research group at DGIST in Korea in a simulation study (Son, et al., 2011) and produced results comparable to those obtained in the on-road environment. The overall protocol consisted of the following:

- welcoming of the participant,
- a brief overview of the experimental procedure,
- review and signing of an informed consent form and other associated participation forms,
- a review of eligibility criteria,
- attachment of physiological sensors
- **in–lab introductory N–back training**

The physiological recording sensors were attached prior to the n-back training and administration of questionnaires to allow participants significant time to adapt to wearing the sensors prior to initiating any actual physiological recordings. The protocol continued with:

- completion of a pre-experimental questionnaire
- baseline physiological recording sitting in a comfortable chair in the intake room
- offering of water and bathroom break
- movement to instrumented vehicle, introduction to vehicle, eye tracking calibration
- **in–vehicle N–back practice**
- approximately 30 minutes of on-road driving (10 minutes to reach highway, 20 minutes on highway before start of assessment period)
- **presentation of N–back secondary cognitive tasks while driving** (details below)

The initial in-lab introductory training was carried out by a Research Associate using a script and support materials that are reproduced in Appendix A and Appendix B. As indicated in the script (Appendix A), the Research Associate started by showing the participant a one page set of instructions (Appendix B) explaining the n-back task. The Research Associate read the instructions out loud while the participant was encouraged to read along. The written hand-out was used so that for part of the first presentation of each task the participant saw a visual representation of the stimulus numbers and the relationship between the presentation of the numbers and how they were expected to respond. This was done deliberately so that participants could see the numbers represented concretely instead of having to rely solely on listening to form a model in their minds of the task expectations. The intent of the overall training protocol is to allow for variations across individuals in learning styles and to maximize the level of understanding of the task across participants.

A review of the task and additional practice trials were later carried out in the parked vehicle before starting the driving portion of the experiment. As indicated in the script (Appendix C), participants were provided with a review of the task instructions for each of the three levels of difficulty. Two, three and four sets of stimuli were then presented as practice for the easy, medium and difficult levels of the task respectively.

For the in-vehicle training (Appendix C), and during the on-road presentation of the task, pre-recorded audio tracks were used to present instructions (Appendix D) and stimulus items (Appendix E). Use of recorded audio ensured that presentation timing and content was identical for all participants. The actual wave files used in the study are available for download on the MIT AgeLab website (http://agelab.mit.edu) to allow interested groups to reproduce the protocol exactly.

Detail on the Timing of the Task & Analysis Periods

Three levels of difficulty of a delayed digit recall task were employed to present drivers with low, moderate and high levels of secondary cognitive workload. The items were presented as recorded auditory stimuli and participants responded verbally. The items consisted of single digits (0-9), presented one at a time, in random order, at an interval of 2.25 seconds between the start of each item presentation.

As can be seen in Table 1, each task period was introduced by recorded instructions lasting 18 seconds that cued the participant as to the version (difficulty level). For example, the easy (0-back) task was introduced with the instructions, "*We are now going to complete a series of scored trials of the zero–back task. Remember that in this task, you are to repeat out loud the number that you just heard. Try to be as accurate as you can be.*" Items were then presented in a block of 4 trials where each trial consisted of 10 randomly ordered stimuli (digits 0-9). Each item (0-9) was presented only once per trial. As noted previously, the spacing between the start of the presentation of each item was 2.25 seconds. There was a brief pause between trials. Each new trial was introduced by the recoding saying, "*Next*". The duration of the total task block for a given difficulty level was 2 minutes. The exact item sets and instructions are reproduced as Appendix E. In the 2008 on-road study, the presentation order for each difficulty level (low, medium or high) was counterbalanced across the sample so that some individuals experienced the low difficulty task first, some the medium, and some the high.

Table 1. Timing Used in 2008 On–Road Study

Period	Duration (min:sec)	Content
Adaptation Period	~30:00	Single Task Driving
Driving (single task)	2:00	Single Task Driving
	0:30 0:18	Separation Interval Task Instructions
Driving + 1st Task	2:00	Four 10 item trials
	2:00 0:18	Recovery Period Task Instructions
Driving + 2nd Task	2:00	Four 10 item trials
	2:00 0:18	Recovery Period Task Instructions
Driving + 3rd Task	2:00	Four 10 item trials
	0:30	Separation Interval
Driving (recovery)	2:00	Single Task Driving

In the 2008 on-road study, a two minute period of "single task" driving prior to the presentation of the first secondary task was used as a reference point for our published analyses on physiological reactivity and visual scanning behavior (Mehler, et al., 2010; Reimer, et al., 2010). It can be noted that a 30 second "separation interval" appears between the end of the initial reference period and the start of the task instructions. We found in our earlier simulation work (Mehler, et al., 2009) that a number of participants showed an increase in physiological arousal just prior to the start of the first cognitive task period. In reviewing the experimental conditions, we developed the impression that in the setting of the simulation laboratory, the Research Associates responsible for monitoring the experiment frequently would tend to shift position in their seats or make other adjustments in anticipation of the initiation of the task period. This likely was noticed by some participants and resulted in modest orienting reactions observed in their physiological recordings. This resulted in our deciding to allow the 30 second separation between the end of the reference period and the start of the instructions to decrease the likelihood that such unintended environmental stimulation was included in the single task reference period.

As indicated in the table, in the primary analyses that we have published to date (Mehler, et al., 2010; Reimer, et al., 2010), a 30 second separation interval was also employed between the end of the last secondary task and the start of the last task driving reference period; this period was labeled as the "recovery" period in these papers. The definition and timing of the initial single task reference period and the recovery reference period are provided for reference purposes only; data analysis periods could certainly be defined differently depending on the intent of a particular analysis.

Training to a Minimal Acceptable Criterion

In these studies, individuals were required to obtain a minimum level of proficiency at the n-back tasks to be included. In the case of the 2008 on-road study, after initial instructions and presentation of practice trials during laboratory n-back training, repetitions of the instructions and practice trails were presented at each task level until participants demonstrated a minimum proficiency of 7 correct responses on the 0 and 1-back (out of 10 & 9 items respectively) and of at least 4 (out of 8) on the 2-back. A maximum of 9 practice trials were allowed for the 2-back. See Appendix A for full details. In the 2008 on-road study, 8 individuals failed to meet the training criterion and were not included in the final analysis set. Unlike the laboratory training, scores from additional practice trials carried out in the parked vehicle were not used as criteria for continuing in the study.

Balancing Engagement and Safety

We believe that it is critically important to develop an appropriate balance between encouraging engagement with the secondary tasks while emphasizing the primacy of safety This is relevant in simulation to encourage behavior that approximates real-world driving as opposed to video game style engagement (Mehler et al., 2009; Reimer, D'Ambrosio, Coughlin, Kafrissen, & Biederman, 2006). The primacy of safety considerations should be explicit for both participants and research staff in on-road studies. The specific instructions reproduced below come directly from the 2008 on-road study (Mehler et al., 2010; Reimer et al. 2010).

In the informed consent form, prospective participants were instructed that, "*During portions of your drive you may be asked to perform several mental tasks. If at any point* *you feel these mental tasks inhibit your ability to drive safely, stop the task and alert the research assistant*." They were also told that, "*During your drive you will receive a small monetary bonus for each correct response to the tasks up to a maximum of $10.*".

These points were covered again later in the informed consent as follows:

"You will have physical control of the vehicle and are responsible for insuring it is operated safely and in accordance with local laws. You are solely responsible for any tickets you may receive while driving the vehicle. Tasks presented during the experiment will be no more demanding than interaction with typical in–vehicle devices or cellular phone conversation. If at any time you feel uncomfortable with the vehicle, tasks presented, or the driving conditions and wish to stop or delay the experiment or tasks you are performing for reason, you are free to do so at any time."

Regarding payment for participation, the informed consent stated that participants would receive a minimum of $50 compensation and that, "*A $10 incentive is dependent on graded performance of the memory tasks you will be performing."*

In actuality, all participants in the 2008 study received $60 compensation regardless of performance.

Prior to practicing the secondary tasks while sitting parked in the vehicle, the following recorded instructions were given to participants:

"During the testing segment, you will receive a small monetary bonus by performing some additional tasks while you are driving. You will hear instructions for these tasks just before you are asked to complete them. These tasks are designed to be difficult for everyone. That is to say, everyone will not be able to answer every question correctly. You will not receive your scores for any of the questions or tasks that are done while driving. You will receive a small amount of additional money for each correct response. Please remember that you will have to pay for traffic violations yourself."

"Again, if you feel that performing these tasks will affect your ability to drive the vehicle safely, please do not perform the task. If at any point during the study you feel uncomfortable with driving the vehicle and your ability to drive safely, please pull off to the side of the road at the nearest safe location."

During the on-road studies referenced in this paper, a research associate was seated in the back of the vehicle to monitor traffic conditions, weather, and the operator's ability to maintain safe control of the vehicle prior to and during all secondary engagements. (The research associate also provided driving instructions and answered questions as needed.) While rare, there were situations where a research assistant directed a participant to safely exit the highway and the research associate drove the vehicle back to the lab. Examples of where this occurred included drivers who appeared sufficiently sleepy to be potentially unsafe and heavy thunderstorms that produced challenging driving conditions.

Selection of research associates who understand the primacy of safety in the research setting, have good judgment, and can project a calm but authoritative presence is important. Research associates are informed that any judgment they make to terminate an on-road experiment will be fully respected and supported as they were the responsible party on-site.

The following notes are printed on the front page of the experimental checklist that research associates complete each session as a means of reinforcing these points:

It is important to remember that when you are involved in data collection with the AgeLab on road vehicle, safety is a priority. Good judgment and attention to detail is key to ensuring the safe collection of good experimental data…

As a research associate involved with this experiment, you are responsible for ensuring that all instructions detailed below are described clearly to the subject and that any questions are answered to the best of your ability. Please make sure to read instructions clearly. If, at any time, the subject has reservations about participating or continuing to participate in this experiment, you should encourage them to withdraw. DO NOT ENCOURAGE PARTICIPATION! …

If, during the course of an experiment, a situation arises where you feel as though the participant's or your safety is compromised, you need to use your best judgment to address the situation. For example, if you feel that the participant is not adequately in control of the vehicle, immediately advise them in a calm manner to pull off to the side of the road at a safe point. Call a contact to report the adverse event. Drive directly back to MIT…

Comment on Task Duration

As detailed earlier in this document, the majority of our work to date has involved the presentation of demand periods consisting of four trials, each trial presenting 10 stimulus items. Each trial interval is approximately 30 seconds in duration, resulting in a total task time of 2 minutes. The 2 minute demand period was selected in part to allow the examination of relatively sustained periods of demand and to have demand periods that were long enough that a wide range of dependent measures (particularly various measures based on variability in driving performance and physiological metrics, i.e. Mehler, Reimer & Wang, 2011) could be accommodated. However, there are no reasons why shorter or longer versions cannot be employed to accommodate variant experimental designs or particular theoretical questions. The lab has in fact done some work with variant length formats and may yet publish some of this data depending on time and resources. The most straightforward approach to varying the duration of the task is to simply add or subtract trials. In other words, use one or more 30 second trials to build overall task duration (i.e. 30, 60, 90, or 120 second long tasks).

Training Material & Items

The appendices that follow provide the training materials and item content of the delayed digit recall task as used to date by the MIT AgeLab. As noted previously, the actual wave files used in the study are available for download on the MIT AgeLab website (http://agelab.mit.edu) to allow interested groups to reproduce the protocol exactly.

Document History:

2011-3	May 11, 2011 – Original release.
2011-3A	June 10, 2011 – Additional background and theoretical consideration of the n-back task, added description of early fixed order protocol, expanded consideration of safety issues, and discussion of task duration considerations.
2011-3B	June 28, 2011 – Added reference to Son et al. (2011) study using n-back protocol.

Suggested citation for this document:

Mehler, B., Reimer, B. & Dusek, J.A. (2011). *MIT AgeLab delayed digit recall task (n–back).* MIT AgeLab White Paper Number 2011–3B. Massachusetts Institute of Technology, Cambridge, MA.

References

Jamson, A.H. & Merat, N. (2005). Surrogate in-vehicle information systems and driver behavior: effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F, 8,* 79-96.

Lenneman, J. K., & Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors, 53(3),* 404-418.

Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138, 6–12.*

Mehler, B., Reimer, B., & Coughlin, J.F. (2010). Physiological reactivity to graded levels of cognitive workload across three age groups: An on-road evaluation. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, Sept. 27-Oct. 1, 2010, 2062-2066.

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single task driving and driving under secondary cognitive workload. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Lake Tahoe*, California,590-597.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25(46),* 46-59.

Reimer, B. (2009). Cognitive task complexity and the impact on drivers' visual tunneling. *Transportation Research Record, 2138*, 13-19.

Reimer, B., D'Ambrosio, L.A., Coughlin, J.F., Kafrissen, M.E., & Biederman, J. (2006). Using self-report data to assess the validity of driving simulation data. *Behavioral Research Methods, 38(2),* 314-324.

Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., & Tan, C. (2009). An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers. *Proceedings of the AutomotiveUI*, Essen, Germany, 115-118.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2010). The impact of systematic variation of cognitive demand on drivers' visual attention across multiple age groups. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, Sept. 27-Oct. 1, 2010, 2052-2056.

Son, J., Mehler, B., Lee, T., Park, Y., Coughlin, J.F., & Reimer, B. (2011). Impact of cognitive workload on physiological arousal and performance in younger and older drivers. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Lake Tahoe, California, 87-94.

Wang, Y., Reimer, B., Mehler, B., Zhang, J., Mehler, A., & Coughlin, J.F. (2010). The impact of repeated cognitive tasks on driving performance and visual attention. *Proceedings of the 3rd International Conference on Applied Human Factors and Ergonomics,* July 17-20, 2010, Miami, Florida.

Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davis (Eds.), *Varieties in attention* (pp. 63-102). London: Academic Press.

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3(2),* 159-177.

Wickens, C.D., & Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. *Human Factors, 30(5),* 599-616.

Zeitlin, L. R. (1993). Subsidiary task measures of driver mental workload: A long-term field study. *Transportation Research Record, 1403*, 23-27.

Zeitlin, L.R. (1995). Estimates of driver mental workload: a long-term field trial of two subsidiary tasks. *Human Factors, 37(3),* 611-621.

Appendix A: In–Laboratory N–Back Training

(Instruction guide used in 2008 study)

Research Associates used the text below in instructing participants. During actual evaluation trials, all instructions were presented from recordings to provide consistency of presentation. For each practice set, the Research Associate writes in the response of the participant and enters the number correct after the word, "Score:".

Part of the experiment will involve performing a set of number tasks. You are going to learn how to perform a few versions of these tasks and practice each with a few trials. This sheet provides an overview of the task.

(Direct the subject's attention to the *N–back Instructions* sheet.)

Please follow along as I explain each version.

The first version is called the **0–back**. During this task, I will read a list of ten single digit numbers. As I read each number, you are to repeat out loud the last number that you've heard. For example, if I were to say the number 3, you would say 3; then if I said 2, you would say 2; then if I said 6, you would say 6, and so on. Try to be as accurate as you can be.

(Point to the appropriate "I say" and "you say" squares on the sheet as you read the above.

I say:	3	2	6	7	1
You say:	3	2	6	7	1

Let's practice with an actual set of numbers:

Score: / 10

7	4	6	8	9	0	5	2	1	3

(If the subject misses more than 1 response, repeat up to four more trials. Write the numbers in the trial above on a separate sheet backwards and then in the same order as they appear alternating up to twice. Present one trial at a time trying to improve the subject's understanding to the point where they respond correctly to seven of ten stimuli.)

The second version of the task is called the **1–back**, which simply means that as I read each list of ten numbers, you are to repeat out loud the number before the last number that you heard. For example, if I said 3, you would say nothing, then if I said 2, you would say 3, then if I said 6, you would say 2, and so on. Try to be as accurate as you can be.

(Point to the appropriate "I say" and "you say" squares on the sheet as you read the above.)

I say:	3	2	6	7	1
You say:	nothing	3	2	6	7

Let's practice with an actual set of numbers:

Score: / 9

9	2	0	7	1	4	6	3	9	8

Let's try that again. Just repeat out loud the number before the last number that you've heard. For example, if I were to say the number 1, you would say nothing, then if I said 2, you would say 1, then if I said 3, you would say 2, and so on. Try to be as accurate as you can be.

Let's practice:

Score: / 9

1	7	3	8	9	0	5	4	6	2

(If the subject misses more than 2 in the last practice trial repeat up to four more trials. Write the numbers in the two trials above on a separate sheet backwards and then in the same order as they appear. Present one trial at a time trying to improve the subject's understanding to the point where they respond correctly to seven of ten stimuli.)

The final version of the task is called the **2–back**, which simply means that as I read each list of ten numbers, you are to repeat out loud the number that was read two numbers ago. For example, if I were to say the number 3, you would say nothing, then if I said the number 2, you would say nothing, then if I said 6, you would say 3, then if I said 7, you would say 2, and so on. Try to be as accurate as you can be.

(Point to the appropriate "I say" and "you say" squares on the sheet as you read the above.)

I say:	3	2	6	7	1
You say:	nothing	nothing	3	2	6

Let's practice with an actual set of numbers:

Score: / 8

5	0	6	7	1	4	2	3	9	8

Let's try another example. Just repeat out loud the number that was read two numbers ago. For example, if I were to say the number 1, you would say nothing, then if I said 2, you would say nothing, then if I said 3, you would say 1, then if I said 4, you would say 2, and so on. Try to be as accurate as you can be.

Let's practice:

Score: / 8

6	5	3	4	7	2	1	8	0	9

Let's try another one. Just repeat out loud the number that was read two numbers ago. For example, if I were to say the number 0, you would say nothing, then if I said 9, you would say nothing, then if I said 1, you would say 0, then if I said 5, you would say 9, and so on. Try to be as accurate as you can be.

Let's practice:

Score: / 8

0	9	1	5	8	2	4	6	3	7

(If subjects miss more than 4 in the last practice trial repeat up to six more trials. Write the numbers in the three trials above on a separate sheet backwards and then in the same order as they appear to form the six trials. Present one trial at a time trying to improve the subjects understanding to the point where they respond correctly to four of ten stimuli)

Good job!

1.	Did the subject complete the 0-back training? (If **no,** the subject is not eligible. Say "These tasks are very difficult to learn. It is not uncommon for people to have difficulty with this part of the experiment, but unfortunately it prevents us from continuing further. I have $50 for you. Thank you for coming in today")	(***YES*** / ***NO***)
	Did the subject complete the 1-back training? (If **no,** the subject is not eligible. Say "These tasks are very difficult to learn. It is not uncommon for people to have difficulty with this part of the experiment, but unfortunately it prevents us from continuing further. I have $50 for you. Thank you for coming in today")	(***YES*** / ***NO***)

	Did the subject complete the 2-back training? Even if the subject didn't complete the 2-back training continue with the subject. (NOTE: Subjects not completing the 2-back training were run through the protocol for data collection purposes but were not considered in the research studies published to date.)	(***YES / NO***)

Summary of training requirements: A minimum proficiency of 7 correct responses on both the 0 and 1-back (out of 10 & 9 items respectively) and of at least 4 (out of 8) on the 2-back. A maximum of 9 practice trials were allowed for the 2-back.

Appendix B: In–Laboratory N–Back Training

(Hand–out participants received in 2008 study)

N–Back Instructions

0–Back

The first version of the task is called the 0-back task, which simply means, that as I read each list of ten numbers, you are to repeat out loud the last number that you've heard. For example, if I were to say the number three, you would say three; then if I said two, you would say two; then if I said six, you would say six, and so on. Try to be as accurate as you can be.

I say:	**3**	**2**	**6**	**7**	**1**
You Say:	**3**	**2**	**6**	**7**	**1**

1–Back

The second version of the task is called the 1-back task, which simply means that as I read each list of ten numbers, you are to repeat out loud the number before the last number that you heard. For example, if I said 3, you would say nothing, then if I said 2, you would say 3, then if I said 6, you would say 2, and so on. Try to be as accurate as you can be.

I say:	**3**	**2**	**6**	**7**	**1**
You Say:		**3**	**2**	**6**	**7**

2–back

The final version of the task is called the 2-back task, which simply means that as I read each list of ten numbers, you are to repeat out loud the number that was read two numbers ago. For example, if I were to say the number 3, you would say nothing, then if I said the number 2, you would say nothing, then if I said 6, you would say 3, if I say 7, you would say 2, and so on. Try to be as accurate as you can be.

I say:	**3**	**2**	**6**	**7**	**1**
You Say:			**3**	**2**	**6**

Appendix C: In–Vehicle N–Back Practice

(Script used for audio recordings in 2008 study)

Text in italic below indicates pre-recorded audio files that played over the instrumented vehicle (or simulator) sound system.

Now we are going to practice the N–back number tasks again. The tasks are the 0–back, 1–back, and the 2–back. We are going to practice them in the order of 0, 1, and 2 for now, but they may be presented in a different order during the actual drive. Each trial will consist of a set of 10 single digit numbers.

The first version of the task is called the 0–back task. As I read each number, you are to repeat out loud the last number that you've heard. For example, if I were to say the number 3, you would say 3; then if I said 2, you would say 2; then if I said 6, you would say 6, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

3	*2*	*6*	*8*	*9*	*0*	*5*	*4*	*1*	*7*

(Pause 5 sec)

Next:

2	*0*	*3*	*1*	*4*	*7*	*6*	*5*	*8*	*9*

The next version of the task is called the 1–back task, which simply means that as I read each list of ten numbers, you are to repeat out loud the number before the last number that you heard. For example, if I were to say 3, you would say nothing, then if I said 2, you would say 3, then if I said 6, you would say 2, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

3	*2*	*6*	*7*	*1*	*4*	*0*	*5*	*9*	*8*

Let's try that again. Just repeat out loud the number before the last number that you've heard. For example, if I were to say the number 1, you would say nothing, then if I said 2, you would say 1, then if I said 3, you would say 2, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

6	*9*	*1*	*7*	*0*	*8*	*4*	*3*	*5*	*2*

(Pause 5 sec)

Next:

2	*0*	*6*	*4*	*3*	*8*	*7*	*1*	*5*	*9*

The final version of the task is called the 2–back task, which simply means that as I read each list of ten numbers you are to repeat out loud the number that was read two numbers ago. For example, if I were to say the number 3, you would say nothing, then if I said the number 2, you would say nothing, then if I said 6, you would say 3, if I said 7, you would say 2, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

3	*2*	*6*	*7*	*1*	*4*	*0*	*5*	*9*	*8*

Let's try another example. Just repeat out loud the number that was read two numbers ago. For example, if I were to say the number 1, you would say nothing, then if I said 2, you would say nothing, then if I said 3, you would say 1, then if I said 4, you would say 2, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

1	*2*	*3*	*4*	*7*	*5*	*6*	*8*	*0*	*9*

Let's try another one. Just repeat out loud the number that was read two numbers ago. For example, if I were to say the number 0, you would say nothing, then if I said 9, you would say nothing, then if I said 1, you would say 0, if I said 5, you would say 9, and so on. Try to be as accurate as you can be.

(Pause 2.25 sec)

Let's practice:

0	*9*	*1*	*5*	*8*	*2*	*4*	*6*	*3*	*7*

(Pause 5 sec)

Next:

4	*5*	*7*	*6*	*3*	*2*	*8*	*9*	*0*	*1*

(end recording n-back_instructions.wav)

Appendix D: In–Vehicle N–Back Secondary Task Instructions

(Script used for audio recordings in 2008 study)

Text in italic below indicates pre-recorded audio files that played over the instrumented vehicle (or simulator) sound system. In the 2008 study, the presentation order of the difficulty level of the N-back task was counterbalanced across subjects so that 1/3rd of the sample was presented with the 0-back first, 1/3rd with the 1-back first, and 1/3rd with the 2-back first. The full set of possible presentation orders (i.e. 0-1-2, 0-2-1, 1-0-2, 1-2-0, etc.) was used across the sample to generate a full counterbalanced design for presentation order. The text below represents the order for the set (0-1-2).

(start recording intro0.wav)

0–back Intro

We are now going to complete a series of scored trials of the 0–back task. Remember that in this task, you are to repeat out loud the number that you just heard. Try to be as accurate as you can be.

(end recording intro0.wav)

(start recording intro1.wav)

1–back Intro

We are now going to complete a series of scored trials of the 1–back task. Remember that in this task, you are to repeat out loud the number before the number that you just heard. Try to be as accurate as you can be.

(end recording intro1.wav)

(start recording intro2.wav)

2–back Intro

We are now going to complete a series of scored trials of the 2–back task. Remember that in this task, you are to repeat out loud the number that you heard two numbers ago. Try to be as accurate as you can be.

(end recording intro2.wav)

Appendix E: In–Vehicle N–Back Task

(Script used for audio recordings in 2008 study)

The boxes below were used by the Research Associate to manually record the type of task and the responses given by the participant. Audio was recorded in the vehicle as well. This double recording method provided redundancy for capturing participant performance.

As can be seen in the structure below, each task consisted of four sets of numbers and was labeled as level (0, 1 or 2) based upon the counterbalanced presentation of the task instructions (Appendix D). Each of these trials consisted of one of the digits 0-9. Each digit is presented once each trail and the order within each trial was originally generated from a random ordering routine. As noted previously, the order of the difficulty level assigned to the first, second and third tasks varied across subjects such that the first block might be presented at the 0, 1 or 2-back level of difficulty. However, the actual items were always presented in the order shown below.

___–back task

(start recording set1.wav)

8	7	4	5	2	3	1	9	6	0

7	3	6	4	0	5	8	1	9	2

2	5	3	4	8	0	7	1	9	6

4	7	0	9	5	3	6	2	1	8

(end recording set1.wav)

(After 2 minutes the system automatically advances to start the 2^{nd} n-back instruction.)

___–back task

(start_recording_set2.wav)

6	5	7	0	1	2	9	8	3	4

9	2	5	3	7	8	1	6	0	4

1	6	7	0	3	9	4	5	2	8

9	0	1	7	3	2	6	8	4	5

(end_recording_set2.wav)

(After 2 minutes the system automatically advances to start the 3rd n-back instruction.)

___–back task

(start_recording_set3.wav)

7	6	0	2	1	3	5	9	4	8

0	4	3	7	5	9	8	1	2	6

3	5	8	1	9	6	0	4	2	7

9	5	1	7	8	3	4	6	0	2

(end_recording_set3.wav)

(Subject was allowed to continue driving uninterrupted for 2.5 minutes.)

2.	Did the subject engage in the entire task (please answer no and provide details if they appeared to stop responding for part of all of the task)	(YES / NO)

About the AgeLab

The AgeLab is a multi-disciplinary research center dedicated to improving quality of life for older adults. Base within the Engineering Systems Division at Massachusetts Institute of Technology, the AgeLab is uniquely suited to translate cutting edge scientific and technological breakthroughs into innovative solutions that help address challenges posed by the world's aging population.

The AgeLab views longevity as an opportunity to innovate – to invent a new definition of quality living throughout the lifespan. AgeLab activities set agendas of government and business, serve as a catalyst for change, and act as platforms to create new ways to remain engaged, connected, independent, and healthy.

Funded by businesses around the world, AgeLab research focuses on transportation, health & wellness, caregiving, longevity planning, shopping, lifelong engagement, and even play. AgeLab research informs the design of new technologies, aids in government policy decisions on the United States and abroad, and educates older adults and their families on important consumer issues.

Contact Information

Agelab
Massachusetts Institute of Technology
77 Massachusetts Avenue
Room E40-291
Cambridge, MA 02139-4307
Phone: (617) 253-0753
Fax: (617) 258-7570
Email: agelab@mit.edu
Website: http://agelab.mit.edu