

# Using machine learning to modify and enhance the daily living questionnaire

Digital Health  
Volume 9: 1–14  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076231169818  
journals.sagepub.com/home/dhj



Peleg Panovka<sup>1</sup>, Yaron Salman<sup>1</sup>, Hagit Hel-Or<sup>1</sup> , Sara Rosenblum<sup>2</sup>,  
Joan Toglia<sup>3</sup>, Naomi Josman<sup>2</sup> and Tal Adami<sup>2,4</sup>

## Abstract

The Daily Living Questionnaire (DLQ) constitutes one of a number of functional cognitive measures, commonly employed in a range of medical and rehabilitation settings. One of the drawbacks of the DLQ is its length which poses an obstacle to conducting efficient and widespread screening of the public and which incurs inaccuracies due to the length and fatigue of the subjects.

**Objective:** This study aims to use Machine Learning (ML) to modify and abridge the DLQ without compromising its fidelity and accuracy.

**Method:** Participants were interviewed in two separate research studies conducted in the United States of America and Israel, and one unified file was created for ML analysis. An ML-based Computerized Adaptive Testing (ML-CAT) algorithm was applied to the DLQ database to create an adaptive testing instrument—with a shortened test form adapted to individual test scores.

**Results:** The ML-CAT approach was shown to reduce the number of tests required on average by 25% per individual when predicting each of the seven DLQ output scores independently and reduce by over 50% when predicting all seven scores concurrently using a single model. These results maintained an accuracy of 95% (5% error) across subject scores. The study pinpoints which DLQ items are more informative in predicting DLQ scores.

**Conclusions:** Applying the ML-CAT model can thus serve to modify, refine and even abridge the current DLQ, thereby enabling wider community screening while also enhancing clinical and research utility.

## Keywords

Functional cognition, daily living questionnaire, machine learning, computerized adaptive testing, ML-CAT

Submission date: 16 October 2022; Acceptance date: 29 March 2023

## Introduction

Functional cognition refers to the use of cognitive processes in daily living activities, namely, the ability to perform daily tasks while using cognitive components such as attention and memory.<sup>1,2</sup> It is defined as “the ability to use and integrate mental and physical skills to accomplish complex everyday activities.”<sup>3</sup> Successful performance of novel or complex activities depends on the ability to efficiently and effectively use and adapt cognitive processes under a wide range of challenges.

Assessment methods for evaluating functional cognition include cognitive or neuropsychological tests, performance-

based assessments and self and informant-report questionnaires. Neuropsychological tests<sup>4</sup> evaluate specific cognitive components and provide baseline profiles, that are norm-referenced yet they do not address the relationship between

<sup>1</sup>Department of Computer Science, University of Haifa, Haifa, Israel

<sup>2</sup>Department of Occupational Therapy, University of Haifa, Haifa, Israel

<sup>3</sup>School of Health and Natural Sciences, Mercy College, Dobbs Ferry, USA

<sup>4</sup>Maccabi Health-Care Services, Tel-Aviv, Israel

### Corresponding author:

Hagit Hel-Or, Department of Computer Science, University of Haifa, Abba Khoushy Ave 199, Haifa 3498838, Israel.

Email: [hagit@cs.haifa.ac.il](mailto:hagit@cs.haifa.ac.il)



cognition and function.<sup>5–7</sup> Performance-based assessments<sup>8</sup> enable direct observation of cognitive processes and the use of strategies during specific functional activities.<sup>9</sup> Examples include the Executive Function (EF) Performance Test<sup>10</sup> and the Kettle test.<sup>11</sup> However, there is, currently, no gold standard performance-based assessment<sup>12</sup> and the number of functional domains that can be evaluated are limited.<sup>13</sup> Questionnaire-based methods are also common in assessing complex everyday activities and functional cognition. Advantages are cost-effectiveness, minimal training requirements, flexible delivery formats (verbally, by phone, electronically) and anonymous administration to a larger population. Self- and informant reports provide a unique, subjective view of the impact of functional cognition on everyday activities in a real-life context.<sup>14</sup>

The Daily Living Questionnaire (DLQ) is a self-report functional cognition instrument that assesses cognitive difficulties in performing higher-level activities including Instrumental Activities of Daily Living (IADLs) (shopping; meal planning), community, work and social participation with the aim of assisting therapists in setting treatment goals.<sup>15</sup> The DLQ is comprised of two parts: Part A focuses on activity and participation, and Part B focuses on everyday cognitive symptoms. See DLQ form in Appendix. Use of self-report questionnaires is recommended,<sup>16</sup> however biases in the form of over- or underestimation of abilities may emerge. In addition, fatigue related to answering a large number of questions may also affect responses, resulting in incomplete answers or low response rates.<sup>16,17</sup> The latter poses an obstacle to conducting efficient and widespread screening of the public for deteriorating functional cognitive abilities. The problem is more acute among individuals with posttrauma, pending medical interventions, long-term COVID and among the elderly community. The goal of this study is to reduce the DLQ questionnaire to a shorter, more efficient assessment.

Machine Learning (ML) algorithms are data-driven and use large datasets to identify patterns in the data and subsequently use these patterns to predict outcomes on novel data. ML is proving to be an effective tool in predicting medical conditions and assisting medical decision-making.<sup>18,19</sup> Computerized adaptive testing (CAT) is an algorithm, based on Item Response Theory (IRT); it adapts or customizes questionnaire items to match individual test-taker ability and predicts the final questionnaire score. The CAT process terminates with a final score when the set number of items has been administered or a precision level has been reached.<sup>20</sup> The CAT approach reduces the average number of items to be administered while maintaining the accuracy of the final test score.<sup>21,22</sup> Additionally, it facilitates more precise assessments that are tuned to the level and abilities of the individual, in comparison to traditional static questionnaires that employ the same questions for all test-takers without taking into account the unique characteristics of individuals.<sup>22,23</sup> In this study we use a novel CAT

algorithm that is based on machine learning and is thus termed ML-CAT. Whereas IRT is a logit-based algorithm, the ML-CAT is based on machine learning and does not place assumptions on the distribution (e.g. normal as in IRT) and the distribution of the data may take any form (restricted to having enough and balanced data in each class). Furthermore, IRT considers the responses of the user an item at a time with older items usually having a smaller effect. The ML-CAT assigns equal importance to all item responses up to that point and may rely heavily on the earlier items in the sequence if indeed they are more predictive.

The aim of this study is to predict the final scores of the DLQ assessment using a reduced set of DLQ questions. The ML-based algorithm, ML-CAT will reduce time and fatigue in completing the DLQ, thereby decreasing patient answer bias to the questionnaire. The ML-CAT can assist in the wide screening of subjects and allow researchers and therapists to collect information in a shorter period of time, with higher accuracy. Our research goals are (1) to develop an efficient ML-CAT scheme with reduced item questions while maintaining the accuracy of the final scores, (2) to determine if a computerized DLQ has the potential to shorten the administration time of the DLQ and (3) to determine if there are key DLQ items that are consistently selected by the ML-CAT.

## Method

### Participants

Correlative studies were conducted between 2007 and 2021 in which participants completed the DLQ and other questionnaires.<sup>15</sup> All studies were approved by the IRB and Ethics committees. One thousand one hundred and eight participants completed the questionnaire. 63.6% were women and mean (SD) age was 55.5 (16.7). Most participants were at a high education level ( $n = 599$ ). Out of the total, 85.9% of the participants were in good health while 14.1% had a medical condition, such as fibromyalgia, multiple sclerosis, or COVID-19, see Table 1.

### Measures

The DLQ is a self-report questionnaire developed to measure functional cognition as reflected through daily activities performance including IADL, at home, work and community as well as social participation. It includes 52 items scored on a 4-point visual analog scale (from 1—no cognitive difficulty to 4—unable to perform) and is divided into two parts: Part A—activity and participation; Part B—cognitive symptoms and impairments. Factor Analysis (FA) revealed seven scores, each calculated by averaging a subset of test items (see Table 2).<sup>15</sup> High internal consistency was found for all 52 items (0.94) and for Part A and Part B (0.87; 0.94, respectively). See Appendix.

**Table 1.** Characteristics of the participants ( $n = 1108$ ).

	%
Gender (M/F)	36.4/63.6
Marital status (married/divorce/single/widower) ( $n = 772$ )	72.2/12.8/8.2/6.9
Education (elementary/high/academic) ( $n = 925$ )	9.4/25.8/67.8
Medical diagnosis (healthy/fibromyalgia/multiple sclerosis/COVID-19)	85.9/7.9/3.2/3.1
Religion (Jewish/Moslem/Druze/Christian/other) ( $n = 904$ )	79.2/9/6.6/4.9/0.3
Residence (city/other) ( $n = 674$ )	75.8/24.2
Living with other (Y/N) ( $n = 750$ )	84.5/15.5
Country (Israel/USA)	86.7/13.3

**Table 2.** DLQ description.

DLQ output label	Name	Number of items (k)	DLQ item numbers
A1	Household tasks	8	1-8
A2	Activities involving language/comprehension	7	9-11, 23-25, 27
A3	Community participation	7	12-14, 17-18, 22, 28
A4	Complex tasks	6	15-16, 19-21, 26
B1	Executive function components	11	29, 31-32, 38-39, 43-46, 51-52
B2	Memory	4	30, 35-37
B3	Executive monitoring	9	33-34, 40-42, 47-50

DLQ: Daily Living Questionnaire.

## Procedures

Patients were recruited across several different research studies and were interviewed using the DLQ.<sup>15</sup> Patients' responses were collected as a database DLQ-DB for use

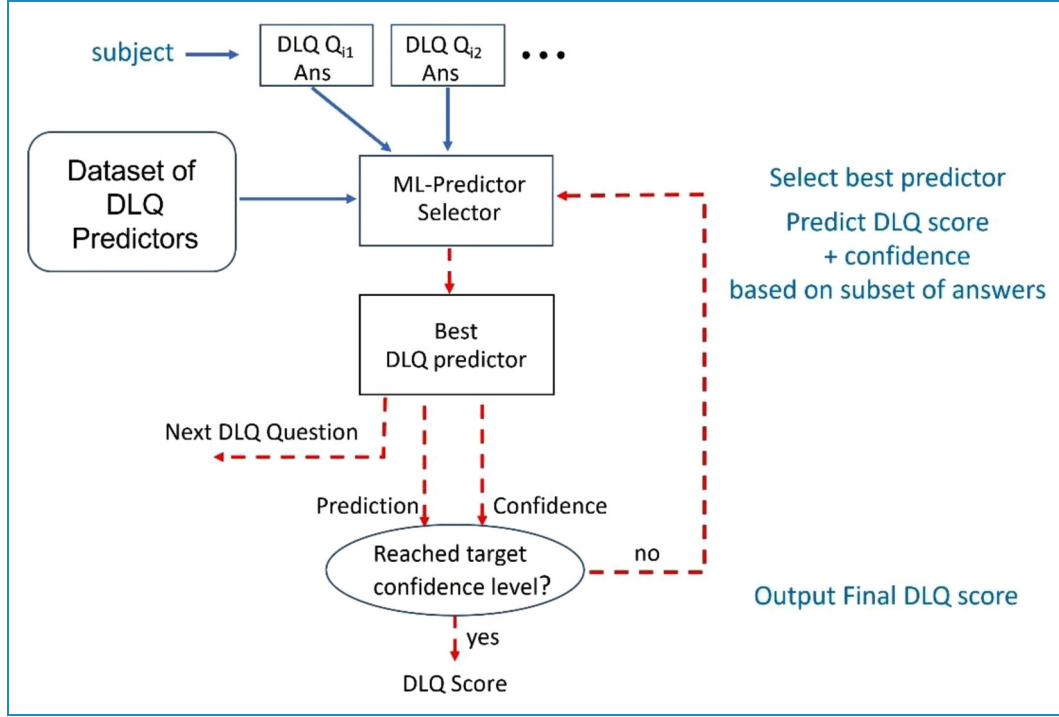
in the ML-CAT method described in this article. Some patients did not answer all questions. Missing data was completed using a Nearest Neighbor approach<sup>24</sup> similar to Hot Deck Imputation,<sup>25</sup> in which each missing score is replaced by the median of the scores of the most similar subjects. Additionally, stratified sampling and augmentation of the data<sup>26</sup> were performed to reduce the bias in scores across subjects in the dataset.

**Machine learning model.** To allow an adaptive testing protocol, we follow a CAT scheme.<sup>27</sup> Our approach is based on a novel machine learning-based CAT algorithm<sup>28</sup> we term ML-CAT. The algorithm iteratively selects the next DLQ question to be administered to the subject based on the subject's responses. Denote the resulting sequence of DLQ questions as  $DLQ_{i1}, DLQ_{i2}, \dots$ . In addition to the next question to be administered, the machine learning component also predicts the final DLQ score<sup>1</sup> at each iteration. The iterative process terminates when the predicted DLQ score reaches a high enough confidence value and the final DLQ score is outputted. A schematic diagram is given in Figure 1. We present two types of ML-CAT algorithms: the first predicts a single DLQ score (thus a separate algorithm is used for each of the seven DLQ-output scores) and the second predicts all seven DLQ-output scores concurrently. Details of the ML-CAT method follow.

Given the DLQ-DB database ( $n = 1108$ ), we build ML predictors to predict the final DLQ score. We train numerous ML predictors each of which receives as input a *subset* of DLQ responses and predicts the final DLQ score together with a confidence measure. Thus, we create a dataset of many such predictors each with a different subset of DLQ responses as input.

Given this dataset of predictors and the DLQ-DB database of testing results, the ML-CAT adaptive testing proceeds as follows. W.l.g, we assume the first DLQ question is given. The following steps are iteratively performed:

1. Assume the subject has already been administered  $k$ -test questions  $DLQ_{i1} \dots DLQ_{ik}$ .
  2. From the dataset of ML predictors, select those predictors that receive as input responses to the set  $DLQ_{i1} \dots DLQ_{ik}$  plus an additional DLQ answer:  $DLQ_{new}$ .
  3. From the database DLQ-DB, select  $M$  samples that are most similar to the subject in their responses to  $DLQ_{i1} \dots DLQ_{ik}$ .
  4. From among the selected predictors (step 2) find the predictor  $ML_{k+1}$  that is most accurate in predicting the final DLQ score on the  $M$  samples selected in step 3.
  5. Set  $DLQ_{new}$  as the next question ( $DLQ_{i(k+1)}$ ) to be administered to the subject.
  6. If the confidence of the prediction of  $ML_{k+1}$  on the  $M$  samples (step 3) is above threshold  $T$ 
    - 6.1 Output the final score prediction of model  $ML_{k+1}$
    - 6.2 Stop program.
- Else:



**Figure 1.** DLQ adaptive testing. Flow diagram describing the adaptive testing scheme (see text).

7. Return to step 1 with the subject's answers to  $DLQ_{i1} \dots DLQ_{ik}, DLQ_{i(k+1)}$ .

The adaptive scheme requires two parameters:  $M$  - the number of samples chosen from the database DLQ-DB and threshold  $T$  which determines the level of confidence required for the final prediction.

**Detailed description of the modeling.** The schema of the adaptive testing detailed above may be implemented using any number of ML models. In Eichler et al.,<sup>28</sup> Random Forest<sup>29</sup> was used as the predictor architecture. In this study, we used linear regression models as the predictors since this sufficed to provide excellent results. Each regressor model predictor is trained by receiving as input a subset of  $k$  DLQ answers,  $DLQ_{i1} \dots DLQ_{ik}$  for each of the  $M$  samples (step 3) together with the final DLQ score  $DLQ_{final}$  for each sample. The regressor model computes the regression variables that minimize the prediction error  $err_k^M$  over all  $M$  samples. The standard mean square error (MSE) was used:

$$err_k^M = MSE = \frac{1}{M} \sum_{i=1}^m (DLQ_{predict_i} - DLQ_{final_i})^2 \quad (1)$$

In addition to the predicted final score, the regression model also outputs the prediction error  $err_k^M$  itself which serves instead of the confidence value described above. Thus,

instead of testing for confidence (step 6), the adaptive testing terminates when the prediction error is below a given accuracy threshold  $T$ . L1 norm (sum of absolute differences) was used to select the  $M$  most similar samples (Step 3).

## Results

The ML-CAT adaptive testing scheme was implemented and tested on the DLQ-DB database of assessment results. The approach was tested for the prediction of each of the seven DLQ output scores (A1, A2, A3, A4, B1, B2, B3) independently as well as predicting all seven scores concurrently.

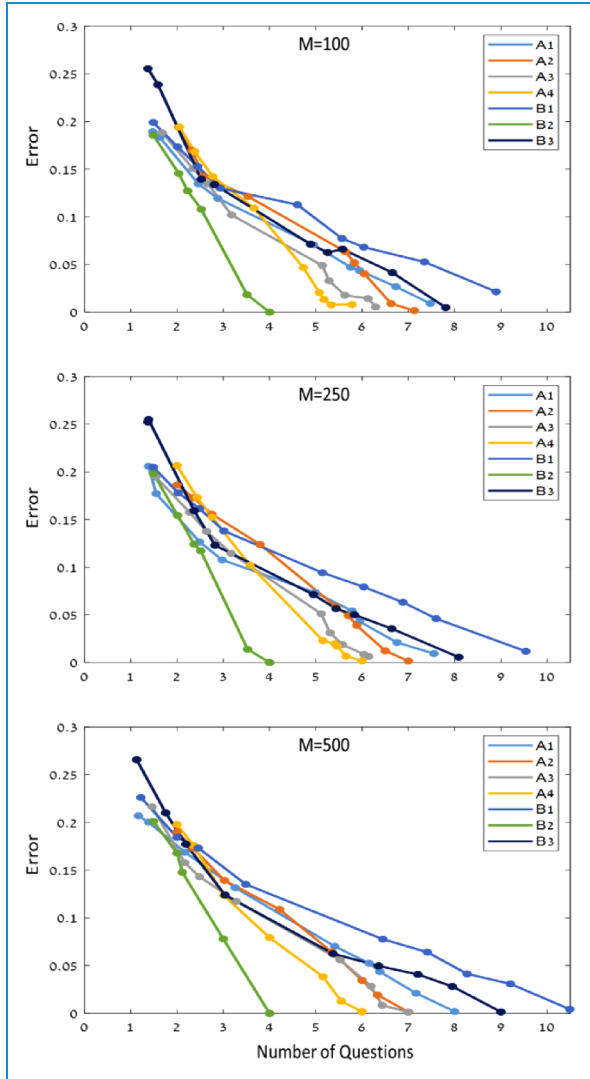
A leave one out<sup>30</sup> scheme was used to assess the adaptive algorithm, where each sample in the DLQ database was used as a subject and the remaining samples were considered as the training set (i.e. as the DLQ-DB database in the algorithm above). The accuracy of score prediction is calculated as the absolute difference between the predicted score and the true score given in the database.

The models are dependent on two hyperparameters: the level of accuracy  $T$  and the number  $M$  of samples of the database, closest to the subject. These parameters affect both the accuracy of score prediction and the number of questions required of each subject.

Figure 2 plots the average error of prediction (mean absolute difference) as a function of the number of DLQ

questions for the seven output scores. For each output score, points represent different values of accuracy ( $T=0.001, 0.003, 0.005, 0.007, 0.01, 0.03, 0.05, 0.07, 0.1$ ). Three plots are shown, each for a different value of  $M$ , the number of samples in the database to train on ( $M=100, 250, 500$ ).

As can be seen in all plots, the error of prediction decreases with increasing number of questions for all output scores, which, in turn, is controlled by the threshold value  $T$ . Predicting some DLQ scores require more questions than others (e.g. B1 vs. A4) for the same level of accuracy. It can also be seen that the number



**Figure 2.** The average error of prediction (mean absolute difference) for the seven output DLQ scores as a function of the number of DLQ questions. For each output score, points are for different values of accuracy ( $T=0.001, 0.003, 0.005, 0.007, 0.01, 0.03, 0.05, 0.07, 0.1$ ). Three plots are shown: Top:  $M=100$ , middle:  $M=250$ , bottom:  $M=500$ .

of questions to reach a specific level of accuracy, varies across  $M$  values, with some DLQ scores requiring more questions than others for higher  $M$  (e.g. A2 vs. B3).

To continue the analysis, we consider the tradeoff between accuracy and number of questions and set  $M=250$  and  $T=0.01$ . We predicted each DLQ output score independently using the ML-CAT model using these values of parameters. Figure 3 shows the average error obtained for each of the DLQ output scores. All errors are smaller than 0.1 MSE with an average error rate of 0.047. Given that the DLQ scores are on a 1–4 scale, these results show high accuracy (low mean absolute error). Figure 4 shows the average number of questions required per DLQ output score to obtain these error rates. These numbers are much lower than the number required by the standard test with savings between 12% and 41% per DLQ score (see Table 3). Figure 5 displays a histogram of the number of times each of the 52 DLQ questions appeared as one of the first three questions in the adaptive testing sequences. Color coding is according to the seven DLQ output scores in the original questionnaire.

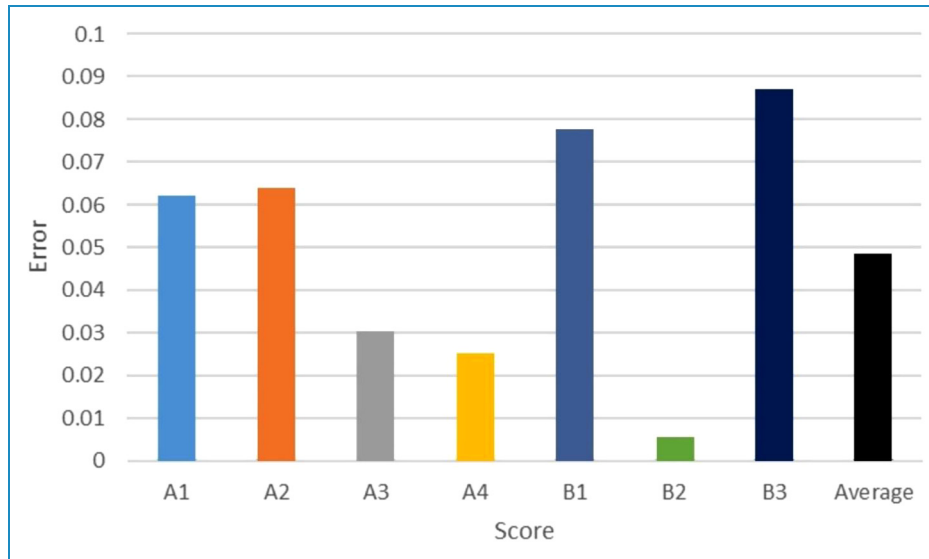
Figure 4 shows that the total number of questions (without repeats) required to predict all DLQ output scores is 38.8 (out of 52) implying a 25.3% reduction. This reduction was obtained when each DLQ output score was predicted independently using a separate ML-CAT process. However further reduction in the number of questions can be obtained by predicting all seven DLQ output scores concurrently, using a single ML-CAT process. Reduction is expected due to correlations between DLQ questions.

### Predicting all DLQ output scores concurrently

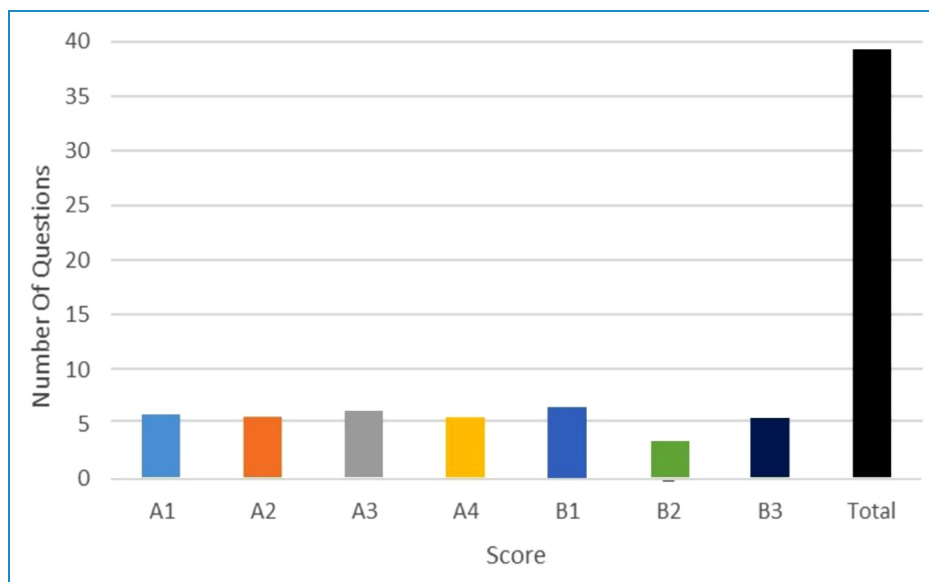
Instead of seven separate ML-CAT models, a single ML-CAT model was built to predict all seven DLQ output scores concurrently. The same  $T$  and  $M$  parameters were used as above. As previously, each regressor model predictor was trained by receiving as input a subset of  $k$  DLQ answers,  $DLQ_{i1} \dots DLQ_{ik}$  for each of the  $M$  samples (step 3) now, however, with the final seven DLQ output scores supplied for each sample. The regressor model computed the regression variables that minimize the prediction error  $errAll_k^M$  over all  $M$  samples by considering the error of all seven DLQ output scores:

$$errAll_k^M = \max_{j=1..7} \left( \frac{1}{M} \sum_{i=1}^m (DLQ_{predict_{ij}} - DLQ_{final_{ij}})^2 \right) \quad (2)$$

where  $DLQ_{predict_{ij}}$  and  $DLQ_{final_{ij}}$  are the predicted value and true value of the  $j$ th DLQ output score for the  $i$ th sample



**Figure 3.** Average error when predicting each output score independently.



**Figure 4.** The average number of questions required to predict each DLQ output score independently with error rates given in Figure 3. Summing the average number of questions required of all labels gives 38.85 questions (compare with the original 52 questions).

(see Equation (1)).

Similar to the above, a leave one out<sup>30</sup> scheme was used to assess the adaptive algorithm. The accuracy of score prediction was calculated as the maximum absolute difference between the predicted and true scores of the subject over all seven output scores.

Figure 6 plots the average error of prediction as a function of the number of DLQ questions for  $M = 50, 250, 500$ . Points on the plot are for different values of accuracy thresholds ( $T = 0.001, 0.003, 0.005, 0.007,$

$0.01, 0.03, 0.05, 0.07, 0.1$ ). The error of prediction decreases with an increasing number of questions, which in turn is controlled by the threshold value  $T$ . However, in comparison to the independent modeling, a higher number of questions were required to reach a level of accuracy when predicting all seven output scores concurrently (compare with Figure 2). This is to be expected as the adaptive algorithm now terminates only when all seven scores are below the threshold (equation (2)). Considering the parameters  $M = 250$  and

T=0.01 as in the models above, an average absolute error of 0.076 (5% error in score value) was obtained with an average of 24.87 questions per subject, implying a 52.2% savings compared to the 25% savings when using independent models. The average error, however, was not uniform across the outcome scores as illustrated in Figure 7. Output scores A4 and B2 show less error and higher accuracy than other output scores. This may be because they require fewer DLQ questions than other output scores and reach higher accuracy before other output scores.

**Table 3.** The average reduction in questions obtained by using the ML-CAT approach compared to the original questionnaire.

DLQ output score	Number of questions original questionnaire	Number of question using ML-CAT	% Saving in questions
A1	8	5.85	27%
A2	7	5.73	18%
A3	7	6.07	13%
A4	6	5.6	6%
B1	11	6.4	41%
B2	4	3.5	12%
B3	9	5.62	38%
Total	52	38.85	25.3%

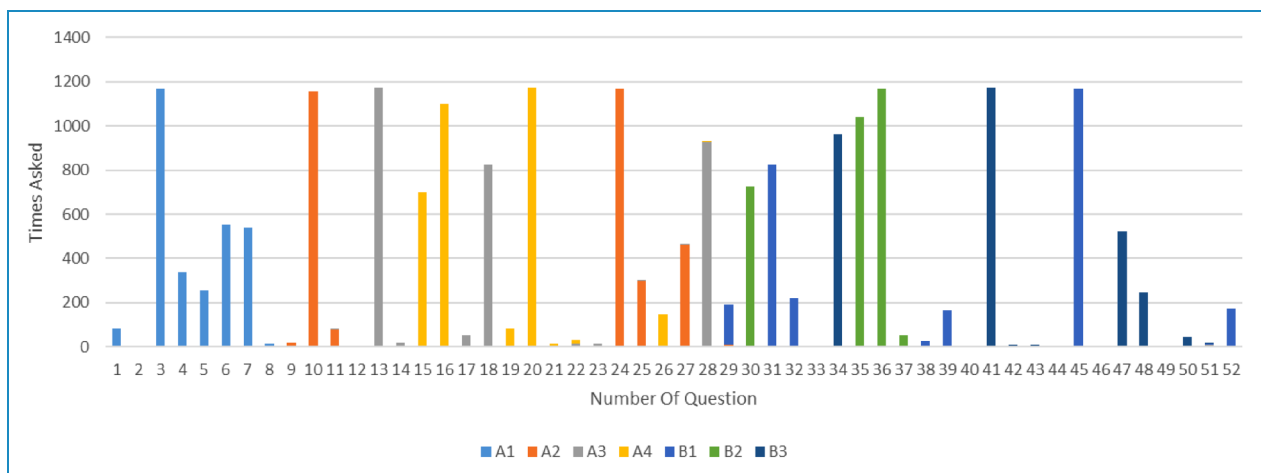
ML-CAT: Machine Learning-based Computerized Adaptive Testing; DLQ: Daily Living Questionnaire.

When applying the ML-CAT approach to DLQ score assessment, every subject is assessed using a different set of questions since questions are selected adaptively per subject. However, it was found that some questions consistently appear among the leading questions for most subjects. Figure 8 displays a histogram of the number of times each of the 52 DLQ questions appeared as one of the first three questions in a subject’s adaptive testing sequence when predicting all outcome measures concurrently. Color coding of the bars are according to the DLQ output scores each question is associated with in the original questionnaire. Questions (13, 20, 31, 45, 51) (Table 4) are frequently selected early in the adaptive test sequence, compared to other DLQ questions. Most likely these questions, although originally associated with only one DLQ output score, are in fact informative to several output scores. They are then chosen by the model, early in the testing process, thus contributing to reducing the number of questions required to reach the set level of accuracy.

### Discussion

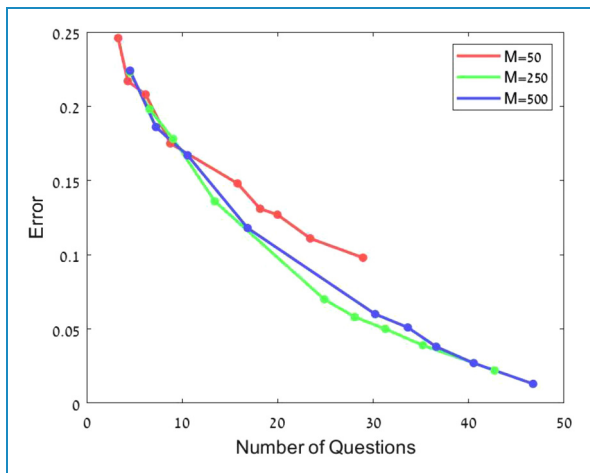
In this study, the ML-CAT algorithm was used to predict DLQ scores in order to adapt and shorten the DLQ assessment. The approach was shown to reduce the number of questions without affecting accuracy, thus shortening administration time.

Items 13, 20, 31, 45, and 51 of the DLQ questionnaire, were most often selected by the ML-CAT early in the test sequence, implying that these items are more informative than other DLQ items. Two of the items are in DLQ part A (activities and participation): social activities with others (item 13) and organizing and managing finances (item 20). Prior FA analysis<sup>15</sup> revealed that item 13 is a component of the community/participation factor (A3),



**Figure 5.** The number of occurrences of each DLQ question as one of the first three questions in the ML-CAT approach when predicting each outcome measure independently. Color coding is according to the seven DLQ output scores in the original questionnaire (n = 2216). ML-CAT: Machine Learning-based Computerized Adaptive Testing; DLQ: Daily Living Questionnaire.

while item 20 is a component of the complex tasks factor (A4). Participation in social activities (item 13) involves the integration of executive functions with other complex cognitive processes that may be impacted by mild cognitive difficulties.<sup>31,32</sup> For example, subtle cognitive deficits can result in misinterpretations of social signals, context, emotions, behaviors, or verbal statements.<sup>33,34</sup> Organizing and managing finances (item 20) have been identified as being sensitive to mild cognitive impairment<sup>35,36</sup> because it requires the integration of different executive function skills. Complex activities such as managing finances and social activities may be

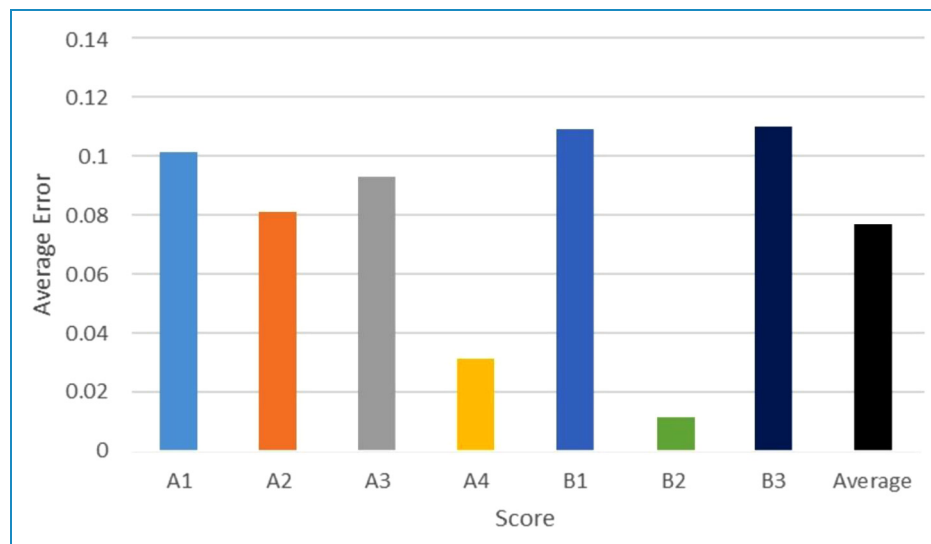


**Figure 6.** Predicting all seven DLQ scores concurrently. The average error of prediction (mean absolute difference) is plotted as a function of the number of DLQ questions for different accuracy thresholds  $T$ . Plots are for  $M = 50, 250, 500$ .

informative in understanding the subject's ability to participate in activities in general and thus is a good predictor of several of the DLQ output scores.

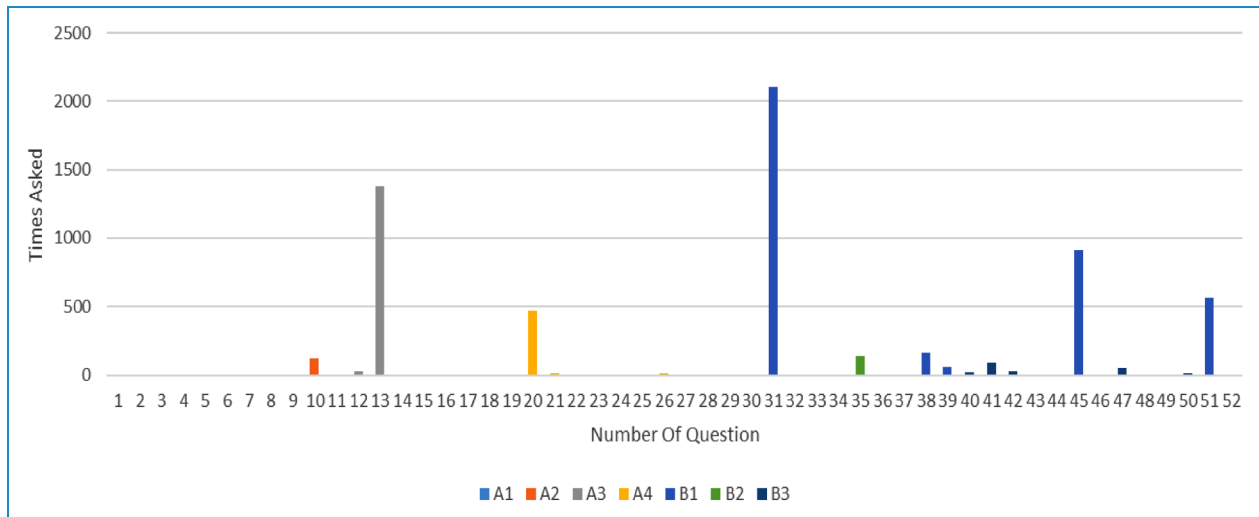
Three of the items are in DLQ part B (cognitive symptoms or impairments): inhibition of irrelevant information (item 31), planning (item 45) and initiation (item 51). Inhibition is the ability to control one's attention, behavior, thoughts and emotions by intentionally preventing the production of an automatic and impulsive response, delaying gratification or resisting distraction.<sup>37,38</sup> Initiation involves generating ideas or plans and beginning activities without procrastination.<sup>38</sup> Planning refers to the capacity to recognize the need for action, define goals, conceptualize a sequence of steps, and choose from alternative methods. Prior factor analysis<sup>15</sup> revealed that all three items are components of the EF factor (B1). The most plausible reason that these three items predict the DLQ scores, is that EF components are considered to be the underlying mechanism for planning and coping with complex cognitive, functional, and social activities.<sup>39-41</sup> Deficient executive function has been found to predict participation and functional outcomes across different health conditions (e.g. Stroke),<sup>42</sup> as well as in healthy populations.<sup>38</sup> EF represents a set of general-purpose control processes that regulate thoughts and behavior,<sup>38</sup> are correlated with skill and engagement in activities,<sup>43</sup> and as such, have a strong impact on all seven factors of the DLQ. Additionally, they are an important aspect when assessing functional-cognition in general, and therefore support the internal validity of the DLQ.

We note a bias in our originally collected data since most participants were healthy. However, the DLQ specifically assesses cognitive difficulties in performing IADL activities,



**Figure 7.** The average error on each output score when predicting all scores concurrently.





**Figure 8.** The number of times each DLQ question appeared as one of the first three questions when predicting all outcome measures concurrently. Color coding of the bars are according to the DLQ output scores each question is associated with in the original questionnaire ( $n = 2216$ ).

DLQ: Daily Living Questionnaire.

**Table 4.** A description of the frequently selected questions by ML-CAT.

DLQ output score	Number of questions original questionnaire	Description
A3	13	Participating in social activities with others
A4	20	Organizing and managing finances
B1	31	Attending to all aspects of a task or situation without missing information
B1	45	Managing multiple step task
B1	51	Taking initiative to start a new activity or project

ML-CAT: Machine Learning-based Computerized Adaptive Testing.

at home, work and in the community as well as social participation. Therefore, unlike an IADL measure that assesses levels of independence, the DLQ would not be expected to show perfect scores in the healthy population (even if indicating high functioning), since there are everyday tasks that present cognitive challenges that may be perceived even by healthy individuals as more mentally challenging. Indeed, not all of our healthy population reported perfect scores (A1—65.3%,

A2—60.1%, A3—62%, A4—46.6%, B1—51.8%, B2—57.1%, B3—64.4%). To ensure any remaining biases the ML algorithms were performed using stratified sampling and augmentation to further reduce any bias effects. Nevertheless, this could have possibly overestimated the accuracy of the short DLQ. We, suggest, as a future study, to test the ML-DLQ on a neurological population. The current study, however, provides a proof-of-concept of an application that can choose the appropriate questions based on the subject's answers, thus, making it easier for clinicians to administer this questionnaire.

## Conclusion

In conclusion, adaptive DLQ testing was implemented using the ML-based ML-CAT, which maximizes test efficiency by selecting the most appropriate items for each individual, without affecting the accuracy of the test score. It was shown that this approach reduces the original questionnaire of 52 items, to less than 25 questions per person on average—a reduction of more than 50%, while maintaining very high accuracy (error rate 0.076 or average 5% error). This abbreviated test may thus facilitate the testing and data collection process, reduce assessment time, and subject fatigue, and thus increase the rate of response. In addition, it may be utilized for wider screening of individuals with cognitive disabilities and other health-related problems in the community.

**Authors' contributions:** Panovka and Salman are principal student authors and contributed equally, Hel-Or advisor. Data collection: Adami, Josman, and Rosenblum. Algorithm

development, machine learning, data analysis and results: Panovka, Salman, and Hel-Or. Writing of manuscript: Panovka, Salman, and Hel-Or. Editing and review of manuscript: Adamit, Josman, and Toglia.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Ethical approval and consent to participate:** All studies were in accordance with the ethical standards of the University of Haifa institutional ethics committee and with Helsinki approval. All participants signed informed consent.

**Funding:** The authors received no financial support for the research, authorship, and/or publication of this article.

**Guarantor:** Hagit Hel-Or.

**Informed Consent:** No images of subjects included in the study. Permission was received from copyright holder to use the questionnaires in this study.

**ORCID iD:** Hagit Hel-Or  <https://orcid.org/0000-0001-8488-0380>

## Note

1. There are seven DLQ output scores. We describe the method for a single output score but will later generalize to multi-output scores. See results below.

## References

- Giles GM, Edwards DF, Baum C, et al. Making functional cognition a professional priority. *Am J Occup Ther* 2020; 74: 7401090010p1–7401090010p6.
- Wesson J and Giles GM. Understating functional cognition. In: Wolf TJ, Edwards DF and Giles GM (eds) *Functional cognition and occupational therapy*. USA: AOTA press, 2019, pp. 7–20.
- Giles GM, Edwards DF, Morrison MT, et al. Screening for functional cognition in post-acute care and the Improving Medicare Post-Acute Care Transformation (IMPACT) act of 2014. *Am J Occup Ther* 2017; 71: 7105090010.
- Marcotte TD, Scott JC, Kamat R, et al. Neuropsychology and the prediction of everyday functioning. In: Marcotte TD and Grants I (eds) *Neuropsychology of everyday functioning*. NY: Guilford, 2010, pp. 5–38.
- Rotenberg S and Maeir A. The cognitive-functional evaluation framework. In: Wolf TJ, Edwards DF and Giles GM (eds) *Functional cognition and occupational therapy*. USA: AOTA press, 2019, pp. 39–46.
- Royall DR, Lauterbach EC, Kaufer D, et al. The cognitive correlates of functional status: A review from the Committee on Research of the American Neuropsychiatric Association. *J Neuropsych Clin Neurosci* 2019; 19: 249–265.
- McAlister C, Schmitter-Edgecombe M and Lamb R. Examination of variables that may affect the relationship between cognition and functional status in individuals with mild cognitive impairment: A meta-analysis. *Arch Clin Neuropsychol* 2016; 31: 123–147.
- Wesson J, Clemson L, Brodaty H, et al. Estimating functional cognition in older adults using observational assessments of task performance in complex everyday activities: A systematic review and evaluation of measurement properties. *Neurosci Biobehav Rev* 2016; 68: 335–360.
- Belchior P, Korner-Bitensky N, Holmes M, et al. Identification and assessment of functional performance in mild cognitive impairment: A survey of occupational therapy practices. *Aust Occup Ther J* 2015; 62: 187–196.
- Baum CM, Morrison T, Hahn M, et al. *Executive function performance test: Test protocol booklet*. St Louis: Washington University School of Medicine, 2003.
- Hartman-Maeir A, Harel H and Katz N. Kettle test – a brief measure of cognitive functional performance: Reliability and validity in stroke rehabilitation. *Am J Occup Ther* 2009; 63: 592–599.
- Sikkes SA, De Lange-de Klerk ES, Pijnenburg YA, et al. A systematic review of instrumental activities of daily living scales in dementia: Room for improvement. *J Neurol Neurosurg Psychiatry* 2009; 80: 7–12.
- Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol* 2012; 34: 11–34.
- Rabin LA, Smart CM, Crane PK, et al. Subjective cognitive decline in older adults: An overview of self-report measures used across 19 international research studies. *J Alzheimers Dis* 2015; 48: S63–S86. PMID: 26402085; PMCID: PMC4617342
- Rosenblum S, Toglia JP and Josman N. Development of Daily Living Questionnaire (DLQ): A factor analysis study. *Open J Occup Ther* 2017; 5: 1–19.
- Patten ML. *Questionnaire research*. 4th ed. London: Routledge, 2016, pp. 1–9.
- Jones S, Murphy F, Edwards M, et al. Doing things differently: Advantages and disadvantages of web questionnaires. *Nurse Res* 2008; 15: 15–26.
- Pourhomayoun M and Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* 2021; 20: 100178.
- Mazza O, Shehory O and Lev N. Machine learning techniques in blood pressure management during the acute phase of ischemic stroke. *Front Neurol* 2022; 14: 743728.
- Yen C, Ho RG, Chen LJ, et al. Development and evaluation of a confidence-weighting computerized adaptive testing. *Edu Technol Soc* 2010; 13: 163–17619. <http://www.jstor.org/stable/jeductechsoci.13.3.163>
- Stocking ML and Lewis C. A new methods of controlling items exposure in computerized adaptive testing. *ETS Res Rep* 1995; 2: i–29.
- Georgiadou E, Triantafillou E and Economides AA. A review of item exposure control strategies for computerized adaptive testing developed from 1983 to 2005. *J Technol Learn Assess* 2004; 5: 1–39. <https://eric.ed.gov/?id=EJ838610>

23. Liu J, Ying Z and Zhang S. A rate function approach to computerized adaptive testing for cognitive diagnosis. *Psychometrika* 2015; 80: 468–490.
24. Yenduri S and Iyengar SS. Performance of imputation methods for incomplete datasets. *Int J Softw Eng Knowl Eng* 2007; 17: 127–152.
25. Andridge RR and Little RJ. A review of hot deck imputation for survey non-response. *Int Stat Rev* 2010; 78: 40–64.
26. Parsons VL. Stratified sampling. In: *Wiley StatsRef: Statistics reference online*. 2014, pp. 1–11.
27. Linden WJ, van der Linden WJ and Glas CA. *Computerized adaptive testing: Theory and practice*. New York, NY: Springer Science & Business Media, 2000.
28. Eichler N, Raz S, Toledano-Shubi A, et al. Automatic and efficient fall risk assessment based on machine learning. *Sensors* 2022; 22: 1557.
29. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
30. Sammut C and Webb GI (eds). *Leave-one-out cross-validation*. In: *Encyclopedia of machine learning*. Boston, MA: Springer Science & Business Media, New York, USA, 2011.
31. Anderson RJ, Simpson AC, Channon S, et al. Social problem solving, social cognition, and mild cognitive impairment in Parkinson's disease. *Behav Neurosci* 2013; 127: 184–192.
32. Rotenberg S and Dawson DR. Characterizing cognition in everyday life of older adults with subjective cognitive decline. *OTJR: Occup Particip Health* 2022; 42: 269–276. <https://doi.org/10.1177/15394492221093310>
33. Toglia J and Foster ER. *The multicontext approach to cognitive rehabilitation: A metacognitive strategy intervention to optimize functional cognition*. Gatekeeper Press, 2022.
34. McDonald S, Gowland A, Randall R, et al. Cognitive factors underpinning poor expressive communication skills after traumatic brain injury: Theory of mind or executive function? *Neuropsychology* 2014; 28: 801–811.
35. Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol* 2012; 34: 11–34.
36. Marshall GA, Zoller AS, Kelly KE, et al. Everyday cognition scale items that best discriminate between and predict progression from clinically normal to mild cognitive impairment. *Curr Alzheimer Res* 2014; 11: 853–861.
37. Miyake A, Friedman NP, Emerson MJ, et al. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognit Psychol* 2000; 41: 49–100.
38. Toglia JM and Katz N. *Cognition, occupation, and participation across the lifespan*. 4th ed. Washington, DC: AOTA press, Washington DC, USA, 2018, pp. 129–142.
39. Cicerone KD, Dahlberg C, Kalmar K, et al. Evidence-based cognitive rehabilitation: Recommendations for clinical practice. *Arch Phys Med Rehabil* 2000; 81: 1596–1615.
40. Liu-Ambrose T, Ahamed Y, Graf P, et al. Older fallers with poor working memory overestimate their postural limits. *Arch Phys Med Rehabil* 2008; 89: 1335–1340.
41. Wood RL and Worthington A. Neurobehavioral abnormalities associated with executive dysfunction after traumatic brain injury. *Front Behav Neurosci* 2017; 11: 195.
42. Adamit T, Maeir A, Ben Assayag E, et al. Impact of first-ever mild stroke on participation at 3 and 6 month post-event: The TABASCO study. *Disabil Rehabil* 2015; 37:667–673.
43. Lewis FD and Horn GJ. Traumatic brain injury: analysis of functional deficits and posthospital rehabilitation outcomes. *J Spec Oper Med* 2013; 13: 56–61. [https://www.neurorestorative.com/assets/sites/7/4\\_Analysis-of-Functional-Deficits.pdf](https://www.neurorestorative.com/assets/sites/7/4_Analysis-of-Functional-Deficits.pdf)

## Appendix—The Daily Living Questionnaire (DLQ)

Name:

Date:

Every day we perform many tasks that require mental effort, also known as cognition. During the course of our lives, we may experience difficulties with these mental efforts or cognitive tasks. The following is a list of activities that you may perform as part of your life.

Please CIRCLE the number which best describes how much mental or cognitive difficulty you generally have doing each of the following activities. If the activity does not apply to you (e.g. you have never done this activity) please circle N/A.

Part 1: How much cognitive or mental difficulty do you have in performing the following tasks?	No mental or cognitive difficulty	Some mental or cognitive difficulty	Much mental or cognitive difficulty	Unable to do	Not applicable (N/A)
1. Getting ready in the morning	1	2	3	4	9
2. Finding items on a crowded shelf or closet	1	2	3	4	9
3. Organizing closets/shelves/draws	1	2	3	4	9
4. Planning and preparing meals	1	2	3	4	9
5. Household tasks (organizing laundry)	1	2	3	4	9
6. Shopping (buying what you need, making decisions, finding items)	1	2	3	4	9
7. Organizing and scheduling own daily activities and errands	1	2	3	4	9
8. Planning/choosing what to wear	1	2	3	4	9
9. Reading newspapers/magazines	1	2	3	4	9
10. Reading books	1	2	3	4	9
11. Searching for information (on internet, library, etc.)	1	2	3	4	9
12. Expressing your thoughts	1	2	3	4	9
13. Following a conversation	1	2	3	4	9
14. Participating in group discussions	1	2	3	4	9
How much mental or cognitive difficulty do you have in performing the following tasks?	No mental or cognitive difficulty	Some mental or cognitive difficulty	Much mental or cognitive difficulty	Unable to do	Not applicable (N/A)
15. Composing a letter or report	1	2	3	4	9
16. Planning social arrangements with family, friends or for children					
17. Participating in social activities with others	1	2	3	4	9
18. Participating in recreational activities, leisure, hobbies	1	2	3	4	9
19. Crossing a busy street	1	2	3	4	9
20. Driving a car	1	2	3	4	9

(continued)

Continued.

Part 1: How much cognitive or mental difficulty do you have in performing the following tasks?	No mental or cognitive difficulty	Some mental or cognitive difficulty	Much mental or cognitive difficulty	Unable to do	Not applicable (N/A)
21. Operating a bank machine	1	2	3	4	9
22. Completing applications and forms	1	2	3	4	9
23. Fixing/repairing things	1	2	3	4	9
24. Finding your way in unfamiliar environments	1	2	3	4	9
25. Math/calculations	1	2	3	4	9
26. Organizing and managing finances	1	2	3	4	9
27. Paying bills	1	2	3	4	9
28. Following written directions	1	2	3	4	9
Part 2: How much difficulty do you have with the following?	No difficulty	Some difficulty	Much difficulty	Unable to do	Not applicable
29. Remembering things you need to do during the day	1	2	3	4	9
30. Keeping track of appointments	1	2	3	4	9
31. Keeping track of where things are	1	2	3	4	9
32. Keeping track of time	1	2	3	4	9
33. Screening out irrelevant background noises or thoughts while engaging in a task	1	2	3	4	9
34. Resuming an activity without difficulty after being interrupted	1	2	3	4	9
35. Prioritizing tasks	1	2	3	4	9
36. Maintaining focus on a task	1	2	3	4	9
37. Switching easily from one task to another	1	2	3	4	9
38. Accomplishing tasks within a reasonable time frame	1	2	3	4	9
39. Responding quickly to situations when necessary	1	2	3	4	9
40. Stopping and starting activities without difficulty	1	2	3	4	9
	1	2	3	4	9

(continued)

Continued.

Part 1: How much cognitive or mental difficulty do you have in performing the following tasks?	No mental or cognitive difficulty	Some mental or cognitive difficulty	Much mental or cognitive difficulty	Unable to do	Not applicable (N/A)
41. Performing daily activities at a normal speed					
42. Understanding new information	1	2	3	4	9
43. Attending to all aspects of a task or situation without missing information	1	2	3	4	9
44. Handling complex tasks that include keeping track of a lot of information at once	1	2	3	4	9
45. Approaching tasks in an organized and efficient way	1	2	3	4	9
How much difficulty do you have with the following?	NO difficulty	SOME difficulty	MUCH difficulty	UNABLE to do	Not Applicable (N/A)
46. Planning and thinking ahead	1	2	3	4	9
47. Seeking out and investigating information when needed	1	2	3	4	9
48. Solving problems without difficulty	1	2	3	4	9
49. Managing multiple step tasks	1	2	3	4	9
50. Adjusting easily to unexpected changes	1	2	3	4	9
51. Taking initiative to start a new activity or project	1	2	3	4	9
52. Learning new factual information	1	2	3	4	9