American Psychological Association

# Neuropsychology

**Manuscript version of**

## Narrative Video Scene Description Task Discriminates Between Levels of Cognitive Impairment in Alzheimer's Disease

Stephanie Reeves, Victoria Williams, Francisco M. Costela, Rocco Palumbo, Olivia Umoren, Mikaila M. Christopher, Deborah Blacker, Russell L. Woods

Funded by:

- Fidelity Biosciences
- National Institutes of Health

# Abstract

The process of interpreting and acting upon the visual environment requires both intact cognitive and visual systems. The narrative description (ND) task, initially developed to detect changes in ecologically relevant visual function in people with impaired vision, is an objective measure of the ability to perceive, understand, and describe a visual scene in a movie clip.

Objective: Since the ND task draws heavily on semantic and working memory ability in addition to basic visual perception, we aimed to assess the discriminative performance of this task across levels of cognitive impairment.

Method: We recruited 56 participants with cognitive status ranging from normal cognition to mild dementia (median age 82, range 66 to 99 years) to watch twenty 30-second video clips and describe the visual content without time constraints. These verbal responses were transcribed and processed to generate ND shared word scores using a "wisdom of the crowd", natural-language processing approach. We compared ND scores across diagnostic groups, and used linear mixed models to examine decrements in task performance.

Results: There was a stepwise decline of ND scores with increasing levels of cognitive impairment. Additional analyses showed that ND performance was highly related to performance on the Montreal Cognitive Assessment (MoCA) and domain-specific neuropsychological tests for semantic fluency and set shifting. Other models demonstrated differences in ND performance related video content between cognitively normal and impaired participants.

Conclusion: The ND test was able to detect decrements in task performance between levels of cognitive impairment and was related to other global neuropsychological measures.

Keywords: Alzheimer's Disease, cognitive impairment, narrative description, assessment, natural language processing

## Key Points

Question: This study asked if cognitive and linguistic deficits could be detected using a task of visual scene comprehension among individuals with cognitive impairment resulting from Alzheimer's Disease.

Findings: Results suggest that individuals with high levels of cognitive impairment performed worse than those with milder forms of cognitive impairment and that the task is associated with other neuropsychological measures.

Importance: These findings confirm a novel way to examine cognitive impairment through a global measure of cognition and linguistic ability.

Next Steps: Future studies might apply this method to remote data collection/monitoring within a clinical trial or observational study for ease of use.

# 1. Introduction

Alzheimer's Disease (AD) is estimated to affect 5.4 million Americans, and by 2050 this figure is expected to rise to 13.8 million. AD is the most common cause of dementia in the U.S.A. and is the sixth leading cause of death nationwide (Alzheimer's Association, 2016). On neuropsychological assessment, early cognitive deficits in AD include impaired episodic memory and language ability, consistent with initial pathological accumulation of amyloid within medial temporal lobe structures (Braak & Braak, 1997; Thal et al., 2000). Prominently observed impairments in episodic memory include deficits in free recall, recognition, and paired-associate learning, suggesting ineffective memory consolidation in this population as information is acquired but quickly forgotten. (Weintraub, Wicklund, & Salmon, 2012). Deficits in language and impaired semantic knowledge are also common in the early stages of AD, as demonstrated by poor performance on tests of object naming, verbal fluency, and word-to-picture-matching (Hodges, Salmon, & Butters, 1992). As AD pathology spreads to include parietal and frontal areas, additional deficits in visuospatial and executive functioning are also noted (Cho et al., 2016; Ossenkoppele et al., 2016).

Despite early and pronounced deficits in language processing in AD, few studies have capitalized on narrative language tasks for informing early AD classification or tracking cognitive decline. Semi-structured interviews used to examine connected speech through open-ended prompts on topics such as career, hobbies, family, and happy memories have been applied to tease apart linguistic differences between subjects with varying levels of dementia (Aramaki, Shikata, Miyabe, & Kinoshita, 2016; Sajjadi, Patterson, Tomek, & Nestor, 2012); however, individual responses are highly variable and there is little standardization in scoring between subjects. Conversely, tasks that involve picture description (narration) prompt more nuanced responses that can be compared to one another through statistical modeling. The most widely used neuropsychological measure that uses verbal descriptions of a visual scene is "The Cookie Theft Picture," which was developed as part of the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983). In this task, individuals are asked to orally describe a simple line drawing that depicts a kitchen in disarray. In a recent review that examined 36 studies of picture description tasks in AD, 25 (69%) used "The Cookie Theft Picture" (Mueller, Hermann, Mecollari, & Turkstra, 2018). Descriptions of Norman Rockwell paintings have also been used to elicit similar natural-language responses (Bayles, Tomoeda, Kaszniak, Stern, & Eagans, 1985; Tomoeda, Bayles, Trosset, Azuma, & McGeagh, 1996). These tasks are able to detect language impairment in AD populations even among those in the earliest stages of the disease, suggesting the potential benefit of narrative description tasks that tap into multiple domains (linguistic, cognitive, and visual) for early detection and monitoring (Forbes-McKay & Venneri, 2005).

Although traditional scoring approaches for picture description tasks utilize a rubric to award points for correctly produced "information units," this approach can be clinically cumbersome to score, and a lack of alternate versions of these tasks weakens validity on repeat assessment. More recently, computer-based natural language processing using data-driven approaches has been applied to characterize task performance more efficiently and objectively. Qualities of speech such as syntactic complexity, repetitiveness, acoustics, content words, and mean word length have been analyzed between groups with and without cognitive impairment, with results that confirm fewer auxiliary verbs, increased pauses, fewer content words and nouns, and significantly shorter words in people with cognitive impairment (Fraser, Meltzer, & Rudzicz, 2016; Kave & Goral, 2016). Increased pronouns are also noted in patients with AD, suggesting empty speech (e.g., nonspecific language, circumlocutions, and verbosity). Other studies have analyzed speech for vocabulary richness and phonetic production using computational approaches for automated assessment (Hernandez-Dominguez, Ratte, Sierra-Martinez, & Roche-Bergua, 2018).

Basic narrative description tasks have proven sensitive in detecting differences between individuals with and without cognitive impairment. But as technology improves, the ways that cognition is assessed will also change, as already evidenced by tablet-embedded voice recorders, digital pens, and advanced machine learning modeling methods (Au, Piers, & Devine, 2017). Narrative description tasks should evolve to reflect this shift in technology to accommodate more dynamic and nuanced scenes. The Cookie Theft picture, among other narration tasks that utilize static images, is adept at evaluating abilities in confrontational naming, verbal fluency, and semantic representations, but not at detecting deficiencies in understanding a conversation, relating facial expressions, connecting scenes across time, or building a narrative based on changing actions. In this way, temporal and spatial relations can only be inferred from a static image as the observer cannot witness and process the evolution of a scene in real time. Such tests can be lengthy, and thus an alternative test that is able to discriminate between levels of cognitive impairment in a naturalistic way that saves clinician assessment time while maintaining sensitivity would be of high clinical utility. We propose that a movie clip narrative description task might be more sensitive than static scene description tasks in discriminating levels of cognitive impairment by requiring participants to link temporal narratives, integrate sequences of actions, and connect relationships among various events. The additional demands of processing dynamic scenes draws on cognitive skills across several neuropsychological domains (visual, linguistic, and working memory) which may offer a more comprehensive summary of cognitive status while also more closely representing real world demands. To our knowledge, there have been no published accounts of a narrative description task that uses movies over static images in people with cognitive impairment.

Our laboratory has developed a task that objectively measures the ability to interpret and report dynamic visual scene information using a natural language processing approach. This method has been previously used to demonstrate decrements in task performance subsequent to visual processing deficits among patients with hemianopia and central vision loss (Costela, Saunders, Kajtezovic, Rose, & Woods, 2018; Costela et al., 2019). Given that describing a visual scene relies on intact working memory and language abilities in addition to visual perception, we thought that this task may offer clinical utility in assessing the cognitive status of patients with progressive neurodegenerative disease. In this study, we aimed to assess the sensitivity of this task in differentiating between levels of cognitive impairment across the AD continuum, and to characterize this task in relation to performance on classic neuropsychological assessments. We hypothesized that: (1) performance on the task would decline across levels of cognitive impairment (intact cognition, impaired cognition not sufficient to meet mild cognitive impairment (MCI) criteria, MCI, and AD); (2) performance on the narrative description task would be closely related to performance on classic neuropsychological tests; (3) ND scores would predict cognitive status to a similar degree as the Clinical Dementia Rating–Sum of Boxes (CDR-SB) and the neuropsychological tests; and (4) attributes of the viewed scenes would affect task performance, and that any effect would vary across levels of cognitive impairment.

# 2. Methods

## 2.1 Participants

Participants were recruited into this study from the longitudinal cohort of the Massachusetts Alzheimer's Disease Research Center (MADRC) at Massachusetts General Hospital (MGH). Individuals in the parent study have been recruited from the Memory Disorders Unit at MGH, from earlier longitudinal studies, and from the community from 2005 on an ongoing basis. Inclusion criteria for the MADRC cohort include age >50 years (with almost all over age 65), ability to attend annual visits, and the availability of a collateral informant. Annual evaluations include the National Alzheimer's Disease Coordinating Center Uniform Data Set (Besser et al., 2018), which includes a clinical evaluation and diagnosis by standard research criteria (Besser et al., 2018), neurological examination, Clinical Dementia Rating (Hughes, Berg, Danziger, Coben, & Martin, 1982), and a standard neuropsychological test battery (see below).

Participants for the present study were recruited from 2015 to 2018. Participants were eligible for the study if they were over age 65, had normal vision with or without corrective lenses, and had a CDR rating of 0, 0.5, or 1.0. We chose to limit the sample to clinical diagnosis of dementia or MCI due to AD based on standard research criteria

(Albert et al., 2011; McKhann et al., 2011) to avoid the oversampling of rare dementias like fronto-temporal dementia and dementia with Lewy Bodies in the overall MADRC sample, but did not require AD biomarkers.

All subjects had binocular visual acuity of 0.30 logMAR (20/40) or better and no known ophthalmic condition that would affect the ability to complete the task (many had undergone successful cataract and intra-ocular lens surgery). As described in Table 1, participants were assigned to the following cognitive status categories based on standard Uniform Data Set (UDS) syndromic diagnostic categories: Normal, Impaired-Not-MCI (i.e., cognitive concerns not matched by objective cognitive deficits or vice versa), MCI, and Dementia (Weintraub et al., 2018).

INSERT TABLE 1 HERE

## 2.2 Cognitive and Clinical Assessment

Cognitive function was assessed at each participant's annual MADRC clinic visit using the standardized Alzheimer's Disease Center UDS - version 3 (Weintraub et al., 2018) consisting of the following subtests: the Montreal Cognitive Assessment (MoCA), Craft Story 21 (immediate and delayed recall), Benson Complex Figure Copy (immediate and delayed recall), Number Span Test (forwards and backwards), Category Fluency (animals and vegetables), Trail Making Test A and B, Multilingual Naming Test (MINT), and Verbal Fluency ("F" and "L") (Slooter et al., 1998). An additional cognitive test, the Free and Cued Selective Reminding Task, was also administered to all participants in the MADRC longitudinal cohort.

The CDR rating was based upon information gathered from interviews conducted by the study physician (neurologist or geriatric psychiatrist) including a direct interview with the participant and written and/or verbal input from the collateral informant following a questionnaire developed and validated at MADRC (Okereke et al., 2011; Okereke et al., 2012). This yielded ratings on each of six categories (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care), each rated on a scale of 0 to 3 in increments of 0.5. Following standard CDR methods, the sum of these categories was used to generate the CDR "sum of boxes" score, and an algorithm was used to determine the global CDR rating (Hughes et al., 1982). Syndromic and etiologic diagnosis was assigned by the study physician following standard research criteria as implemented in the UDS protocol (Morris et al., 2006).

## 2.3 Narrative Description Task

The narrative description (ND) task is an objective measure of the ability to perceive, understand, and describe a visual scene, using natural-language descriptions made by an observer (Saunders, Bex, Rose, & Woods, 2014). Participants were asked to view a set of twenty 30-second movie clips, each in a different randomized order, and describe what they observed, without time constraints, into an audio recorder. Audio content from the clips was not available, and thus responses were restricted to descriptions of visual content alone. The prompt was: "Describe this clip as if you were describing it to someone who has not seen it before," and then, "Are there any other details about the clip that you might want to mention?" After describing the clip, participants were asked if they had seen the movie ("Have you seen this movie before?"), and if not, were then asked if they would be interested in watching the movie on a scale of 1 (not interested) to 5 (very interested). All subjects were also asked if they liked the movie clip ("How much did you like the clip?") on a scale from 1 (not at all) to 5 (very much). This process was repeated for all 20 video clips. Study participants viewed the visual content on a 27-inch screen (display diagonal 73 degrees visual angle) seated from about 50cm away.

## 2.4 Video Clips and Scoring

We used 20 video clips drawn from a publicly available (shared) database of 200 30-second "Hollywood" video clips (movies, television, and documentaries), chosen to represent a range of genres and types of activities (Costela & Woods, 2019). In that dataset, the extracted 30-second clips were chosen to contain a cohesive "scene" (typically with only one setting) without much fast action. As a consequence, the video clips had fewer scene cuts (average of nine cuts per clip in the database) than typical movies (Costela & Woods, 2019). The genres included dramas (e.g., *Julie & Julia*), nature documentaries (e.g., *March of the Penguins*), cartoons (e.g., *Shrek*), and action/adventures (e.g., *The Stepfather*). Clips had been rated by two observers, who did not participate in our study, for the importance for clip understanding of faces, human figures, man-made objects, and nature on a scale from 0 to 5, with 0 being absent and 5 being always important (Costela & Woods, 2018).

The procedures for scoring the transcriptions of the raw audio files are described in detail in Saunders et al. (2014). Essentially, the transcriptions were objectively scored for relevant content using an automated "wisdom of the crowd" (Surowiecki, 2004) approach that generated a shared word score. In generating that score, each response was compared to reference control responses of the same video clip from a database of 126 individuals consisting of 66 crowd-sourced plus 60 lab-sourced participants. These participants were 22-85 years old, 76 were female (50 male), and 51% had at least a bachelor's degree; see Saunders et al. (2013) for more information about this group. A comparison between the lab-sourced reference sample (n=60) and the participants with normal cognition recruited as part of the present study (n=14) is presented in section 3.2. The number of words (after removing stop words such as "um", "ah", "is" and "the") shared by each pair of responses (one from the control database and one "new" response), disregarding repeated instances of the word in either response, produced a shared-word count for each pair of responses. The ND score for each video clip for each study participant was the average of the shared-word counts from the paired comparisons with each of the responses from the response database for the same clip. For more details on the video clips and the scoring of natural-language descriptions, see Saunders et al. (2014).

## 2.5 Statistical Analyses

Statistical analyses were conducted using Stata/IC 14 (StataCorp, College Station, TX, USA). All variables were assessed for normality and the presence of outliers prior to statistical analysis. To determine the effect of cognitive status on the narrative description task, we first examined raw data (shown in Table 2) and then used a linear mixed model that accounted for age, education, and gender as covariates, and subject and clip as fully crossed random effects. We examined the relationships between average ND scores and CDR-SB, MoCA, and the 14 neuropsychological subtests with Spearman correlations. We used additional mixed models to predict cognitive status, and to examine the effects of clip characteristics and clip perception on ND scores. Mixed models that make use of all of the data (and not a composite or average score) have advantages such that: (1) the model accounts for differences in difficulty between stimuli (i.e., if everyone sees every stimulus, then an average is a fair representation, but if some subjects only see a subset, then an average is not fair); (2) the model accounts for differences in variability between subjects (this is to do with the "reliability" of each subject; a subject who is more reliable gets weighted a bit more strongly in the model than a subject whose data varies a lot between stimuli); and (3) the model has greater statistical power and more degrees of freedom with all of the data included in the model. We accepted $p \leq 0.01$ as statistically significant, and report terms with $0.10 \geq p > 0.01$ as trends.

# 3. Results

## 3.1 Participant Characteristics

A total of 56 individuals (24 males and 32 females, aged 66-99 years) participated in the study. Of those, 14 participants were classified as cognitively normal, 19 were classified as impaired-not-MCI, 15 were classified as MCI, and 8 were classified as dementia. Demographic information is provided in Table 2**.** Most groups were majority female, and although there was a high proportion of females among the cognitively normal group, there were no significant differences between the groups in distribution ($X^2$ = 4.13, p=0.25). The education level of the sample was generally high (more than 16 years) and did not differ across the groups ($X^2$ = 5.99, p=0.74). As expected, greater age was associated with worsening syndromic diagnosis ($X^2$ = 6.42, p=0.01). The sample self-identified as 87% white, 5% black, and 2% Asian, and there was a somewhat higher proportion of Asian individuals in the dementia group than in the others, though the difference was not statistically significant ($F_{3,52}$=2.32, p=0.09). The ND test took, on average, 28.6 minutes to complete (median 27 minutes, range of 19 to 70) and this did not differ across the four groups ($F_{3,50}$=0.95, p=0.42). All subjects saw 20 video clips, except for eight subjects who provided responses to between 16 and 19 clips.

INSERT TABLE 2 HERE

## 3.2 Comparison between reference group and cognitively normal group

Since the ND method had previously only been applied to samples that were cognitively normal, we ensured that the participants with normal cognition in this study (n = 14) were comparable to the 60 participants from the reference group (used for the processing of ND scores) by running a linear mixed model that compared the two samples. There was no difference in ND scores between the two groups (z=0.69, p=0.49) when corrected for age, gender and education. ND score decreased with age (with a modeled difference of 0.29 shared words per decade; z=3.86, p<0.001) across the two groups equally (z=0.17, p=0.86). There was no significant effect of gender in the reference group (z=0.85, p=0.39) (Costela et al., 2019), but among the participants with normal cognition in our study, males had lower ND scores (1.8 shared words, z=2.55, p=0.01). There was no effect of education in either group (z=1.54, p=0.12).

## 3.3 ND task discriminates across levels of cognitive impairment

We hypothesized that individuals with cognitive impairment would perform worse on the ND task compared to those without cognitive impairment, and that this effect would enlarge as the level of impairment increased. Average ND scores for each group are presented in Table 2 and Figure 1a, depicting decrements in task performance across worsening levels of cognitive impairment. To further examine this effect of cognitive status on ND score, we used a linear mixed model with “participant” and “video clip” as fully-crossed random factors, and age, gender, and education as covariates. As expected, the average adjusted ND scores decreased with a step-wise decline across worsening levels of cognitive impairment (Figure 1b). Compared to participants with normal cognition, those with dementia scored 2.5 shared words worse ($z = -4.60$, $p<0.001$); those with MCI scored 1.3 shared words worse ($z = -2.88$, $p=0.004$); and those with impaired cognition but not meeting MCI criteria (“Impaired-not-MCI”) scored 0.8 shared words worse ($z = -2.01$, $p=0.04$). There was a difference in ND scores between the MCI and dementia groups ($X^2 = 5.59$, $p=0.02$), but not between the impaired-not-MCI and MCI groups ($X^2 = 1.18$, $p=0.28$). We also conducted this analysis with the inclusion of the factor “seen movie before” with the expectation that the variable might be a confounding factor, but found that there were no substantive differences in the effects of cognitive status on ND scores; see section 3.7 for further details.

INSERT FIGURE 1 HERE

## 3.4 ND Performance Associates with Neuropsychological Assessment

We asked if performance on the ND task was related to performance on the CDR-SB functional measure and on classic neuropsychological measures. Thus, we ran pairwise Spearman correlations between the ND task and the CDR-SB and both global and domain-specific neuropsychological tests. As the Spearman correlation is a rank-order correlation, it does not minimize residuals, and thus is not affected by differing test ranges. As shown in Table 3, ND was most strongly related to general measures of global cognitive status: the MoCA (rho = 0.71, $p<0.001$) and CDR-SB (rho = -0.45, $p<0.001$). The ND task was also associated with measures of linguistic ability: MINT (rho = 0.52, p<0.001), category fluency (rho = 0.63, p<0.001), and letter fluency (rho = 0.52, p<0.001); measures of executive functioning: Trail Making Test (B) (rho = -0.56, p< 0.001) and (B-A) (rho = -0.60, p< 0.001); and measures of memory: Craft Story 21 Delayed (rho = 0.47, p<0.001), Benson Figure Delayed (rho = 0.50, p<0.001) and Benson Figure Percent Retention (rho = 0.41, p=0.001).

INSERT TABLE 3 HERE

### 3.5 ND Scores Predict Cognitive Impairment Status

As the ND test taps into multiple domains to evaluate cognition, we hypothesized that ND scores would be able to predict diagnostic category (cognitive status). To examine this, we employed a backwards stepwise ordinal logistic regression to predict diagnostic category (cognitive status) from neuropsychological measures and CDR-SB. All neuropsychological measures were univariate related to cognitive status (p<0.03), save for Number Span Backwards (p=0.12) and Craft Percent Retention (p=0.06), thus those two metrics were omitted from the initial predictors in the stepwise logistic regression. To reduce potential multi-collinearity problems, Craft Story Immediate, Benson Complex Figure Copy, Benson Complex Figure Delay, Trail Making Test part B were not included. Age, gender, and education were included in the model as covariates. The full model began with all variables (except those noted above), and systematically removed, one at a time, those that did not significantly account for diagnostic category variance (p>0.10). In the final model, ND score (z=2.69, p=0.009) and CDR-SB (z=4.69, p<0.001) were the only remaining predictors of cognitive status in our sample. The ND task and CDR-SB were better predictors of cognitive status than the many other domain-specific neuropsychological tests. CDR-SB is a measure of functional status and is operationally linked to the determination of diagnostic groups, and thus was expected to be highly predictive. It is possible that among the neuropsychological measures, the ND test was highly predictive because it tapped additional cognitive domains beyond isolated abilities related to semantic and working memory ability. A summary of statistical tests comparing the regression models is provided in Table 4.

INSERT TABLE 4 HERE

### 3.6 Video Clip Attributes Impact ND Score

To examine how certain characteristics of video clips affected ND score, we examined the effects of these features on understanding the clip: man-made objects, human faces, human figures, and nature (importance, on a scale from 0 to 5). We hypothesized that these features may have an effect on ND scores and that this effect might differ between participants with normal cognition and those with impaired cognition. First, we examined participants with normal cognition. Figure 2 depicts ND scores and regression lines for the range of nature and object rankings. With a high importance of nature, ND scores were low; alternatively, with a high importance of man-made objects, ND scores were high. To compare the effects of all four importance features between participants with and without cognitive impairment, we fit a linear mixed model with interaction effects between cognitive impairment presence (a binary variable) and each clip characteristic with "participant" and "video clip" as fully-crossed random factors. In the cognitively normal group, as shown in Figure 3, ND scores increased by 0.41 shared words per importance unit of objects ($z = 4.01$, $p<0.001$), scores decreased by 0.32 shared words per importance unit of nature ($z = 2.82$, $p=0.005$), and decreased by 0.28 shared words per importance unit of faces (z = 2.20, p=0.03). Next, we examined participants with impaired cognition. We found no factor effects of the importance of nature (z=0.99, p=0.32) or faces (z=1.23, p=0.22) in the cognitively impaired group (Figure 3), however, the importance of man-made objects increased scores by 0.26 shared words per importance unit (z=2.83, p=0.005). There were no significant effects of human figures on ND scores in either the cognitively normal group (z = 0.17, p=0.86) or the cognitively impaired group (z=0.01, p=0.99).

INSERT FIGURE 2 ABOUT HERE

Previously, studies have shown that abilities related to object processing (Laatu, Revonsuo, Jaykka, Portin, & Rinne, 2003; Tippett, Blackwood, & Farah, 2003) and semantic fluency (Verma & Howard, 2012) are affected in cognitively impaired populations, thus, we expected that the effects of video clip characteristics that were found in the cognitively normal group might differ from those found in the group with impaired cognition. Thus, we examined the interaction effects from the linear mixed model described above. When the importance of nature was high, those with impaired cognition performed better on the ND task than participants in the cognitively normal group (z=3.31, p<0.001), although their scores were still low (negative values in Figure 3). Conversely, when the importance of man-made objects was high, individuals with impaired cognition performed worse on the ND task than those in the cognitively normal group (z=2.47, p=0.01), although their scores were still high (positive values in Figure 3). Lastly, the effect of faces was also different from that found in the cognitively normal group, albeit, weakly (z=1.85, p=0.06). Average marginal effects of ND scores for each of the four feature categories are shown in Figure 3.

INSERT **FIGURE 3** ABOUT HERE

### 3.7 Attitude Toward Video Clip Affects ND Score

There is some literature that suggests a person's attitude toward a visual stimulus may affect: (1) their perception of its image quality, and (2) their perception of the stimulus itself (Kortum & Sullivan, 2010; Pozueco et al., 2017; Sullivan, Pratt, & Kortum, 2008). For this reason, we asked participants to subjectively rate their experience of each video clip by asking them three questions. The intent of the questions was to examine whether having seen a movie before, having an interest in a movie, or liking a movie had any effect on the ability to describe it (ND scores). Summary data are shown in Table 5. A linear mixed model was used to determine the effect of participant experience on scores by group with participant and clip included as random factors. Unexpectedly, in the cognitively normal group, having previously seen the movie decreased scores by 0.7 shared words ($z = -3.11$, $p=0.002$), and this effect did not significantly differ between cognitively impaired and normal groups ($p>0.15$). Having an interest in a movie (if they had not seen the movie) did not have an effect on ND scores, except for those in the impaired-not-MCI group who scored 0.2 shared words lower than the cognitively normal group ($z=2.24$, $p=0.025$). Liking the movie had no global effects on ND scores ($p\geq0.40$), except for participants with dementia who scored 0.3 shared words higher than the cognitively normal group ($z = 2.09$, $p=0.037$). Since the participants themselves made these subjective ratings, many of whom had cognitive deficits, episodic memory deficits may have interfered with recollection accuracy among those with cognitive impairment.

## 4. Discussion

Understanding and acting upon the visual environment requires three basic functions: 1) intact visual processing, 2) intact cognitive processing, and 3) intact linguistic or motor processing (to interact with or respond to the visual world). We assessed visual, cognitive, and linguistic function through a narrative description task that uses natural language processing to produce objective scores of content understanding and reporting. We have used this narrative description (ND) method in the past to examine decrements in vision impairments such as central vision loss (Costela et al., 2019) and hemianopia (Costela et al., 2018). In this study, we evaluated whether this task and analysis method could be applied in a novel population to discriminate between levels of cognitive impairment in patients with Alzheimer's Disease. We hypothesized that the ND scores of people with cognitive impairment would be lower than those of the cognitively normal group. As expected, participants with worse levels of syndromic diagnosis had lower ND scores, demonstrating that the ND method can be applied to populations with cognitive impairment and that it can detect differences between groups with different levels of cognitive impairment.

Since the ND task was able to discriminate between different levels of cognitive impairment, we asked if performance on our task was similar to performance on other neuropsychological tests, both global tests and those in related cognitive domains. As expected, the ND task was related to individual neuropsychological tests of confrontational naming (MINT), semantic fluency (category fluency), and aspects of executive functioning (Trail Making Test part B minus part A). Furthermore, we found that ND scores were highly related to the MoCA and CDR-SB, commonly used measures of global cognitive function and of global functional status; and that the ND task and the CDR-SB were the best predictors of cognitive status even above the many domain-specific neuropsychological measures. The relationship between ND scores and these global tests indicates that ND is able to function as a global measure of cognitive status by tapping multiple domains, with similar sensitivity to more standard screening measures such as the CDR-SB and MoCA. While neuropsychological assessment tests identify isolated domains of cognitive impairment, the ND task integrates performance across multiple domains and provides a more complete picture of global impairment.

We found that certain attributes or characteristics of video clips affected ND scores. In our study, high importance of man-made objects increased ND scores in participants with normal cognition, presumably since those clips contain interesting visual features to describe. The presence of nature was associated with lower scores in cognitively normal participants, presumably due to the limited details to describe (e.g., there is only so much to say about penguins walking across a sheet of ice). However, that does not explain the lower ND scores for faces, for which there would seem to be much to describe. In a separate, larger sample (n=63) reported previously (Saunders et al., 2014; Saunders et al., 2013), lower ND scores were found with increasing nature importance (-0.26 shared words/level, $z=3.20$, $p<0.001$), but no effect of face or object importance was found ($p>0.15$). Thus, the effects of face importance may be dubious. Interestingly, the effects of video content on ND scores were substantially muted by cognitive impairment. Unlike the cognitively normal group, the ND scores (quality of descriptions) of subjects with impaired cognition did not vary with video content, so the subjects were saying almost as much about “nature” videos as they were about “objects” videos.

That being said, within our sample, the effects of feature importance (for understanding the video clip) were different between those with cognitive impairment as compared to subjects with normal cognition. As shown in Figure 3, when there was a high importance of man-made objects, individuals with impaired cognition experienced a decrease in ND scores as compared to those with normal cognition. These observations add to the growing literature that suggests difficulties in object processing in AD. Previous studies have shown impairments in visual object recognition (Laatu et al., 2003) and in basic shape processing related to semantic-lexical impairments (Tippett et al., 2003). These issues relate to broader deficits in confrontational naming, semantic memory, and verbal fluency, as patients with AD have greater difficulty relating words to their meanings and objects to their names (Verma & Howard, 2012). Conversely, when there was a high importance of nature, participants with impaired cognition performed better (relatively) on the ND task as compared to those with normal cognition. Some studies have reported general benefits for patients with AD to spend time outdoors in nature (Brawley, 2007), and it is possible that even simply observing nature (from a television) could mimic some of those effects. Other preferences for natural aesthetics in paintings and visual scenes have been documented in AD (Graham, Stockinger, & Leder, 2013), as has increased recollection for animate words (as opposed to inanimate or non-living) (Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013), giving further support for our observation that nature scenes produced increased ND scores in participants with cognitive impairments.

There were some limitations of this study including that the size of the sample was small (n=56), education levels were higher than the national average, and 87% of the participants were white. Given that there are effects of education and cultural differences on perceptual and linguistic processes (Rule, Freeman, & Ambady, 2013), our findings may not be generalizable to the greater population, and future work should aim to implement the ND test in samples with greater educational and ethno-racial diversity. Furthermore, our diagnoses of cognitive impairment due to AD were based on clinical measures rather than AD biomarkers, and the findings must be interpreted within this limitation. Although we restricted the sample of patients with dementia to those with a clinical diagnosis of AD, we make no claims that the ND is related to Alzheimer's disease per se, and indeed our expectation was that the test correlates with syndromic diagnosis (dementia, MCI, etc.), not with underlying pathology.

While the ND task has not yet been applied to contexts outside of the lab, given its ease of administration, future studies might consider it as a suitable measure for longitudinal in-home testing or remote data collection/monitoring within a clinical trial or observational study. One advantage of the ND task is that it is a relatively short: the ND task took, on average, 28 minutes to complete, and did not take significantly longer for the cognitively impaired groups. In fact, those with dementia took the shortest amount of time to complete the test (mean 25 minutes, SD 4.3 minutes). Other advantages of the ND task include its ecological validity, given that the task is akin to real world demands by requiring participants to link scenes temporally, interpret naturalistic actions, and perceive interpersonal relationships. Furthermore, the ND task is an engaging task from which the participant may obtain a small level of enjoyment. Thus, we expect that it would not be difficult to ask patients or research participants to perform the ND task on multiple occasions. No clinical expertise is required to administer the ND test, and in-home delivery could allow close monitoring of cognitive status for medical care and for interventions, reducing the number of in-office visits (thereby reducing costs) or providing alerts of an alteration of status. The ND task adopts an automated and naturalistic approach to assessment using narrative description of a visual scene and in this way is a promising remote monitoring or screening tool that does not require face-to-face administration. Future work should aim to empirically assess the psychometric properties of the ND task as a cognitive screening tool (including diagnostic specificity and sensitivity) in larger and independent samples, and its ability to detect within-subject change in cognitive status. Other work should examine the association between the ND task and existing picture description tasks for further validation of this task. In conclusion, the ND task is a naturalistic test simulating real world demands that serves as a global cognition measure that can differentiate between levels of cognitive impairment.

**Acknowledgements:**

The study was supported in part by National Institutes of Health grants P30AG062421 and P50AG005134 to Massachusetts General Hospital and P30EYEY003790 to the Schepens Eye Research Institute, and by Fidelity Biosciences, through a grant to the Harvard Neurodiscovery Center.

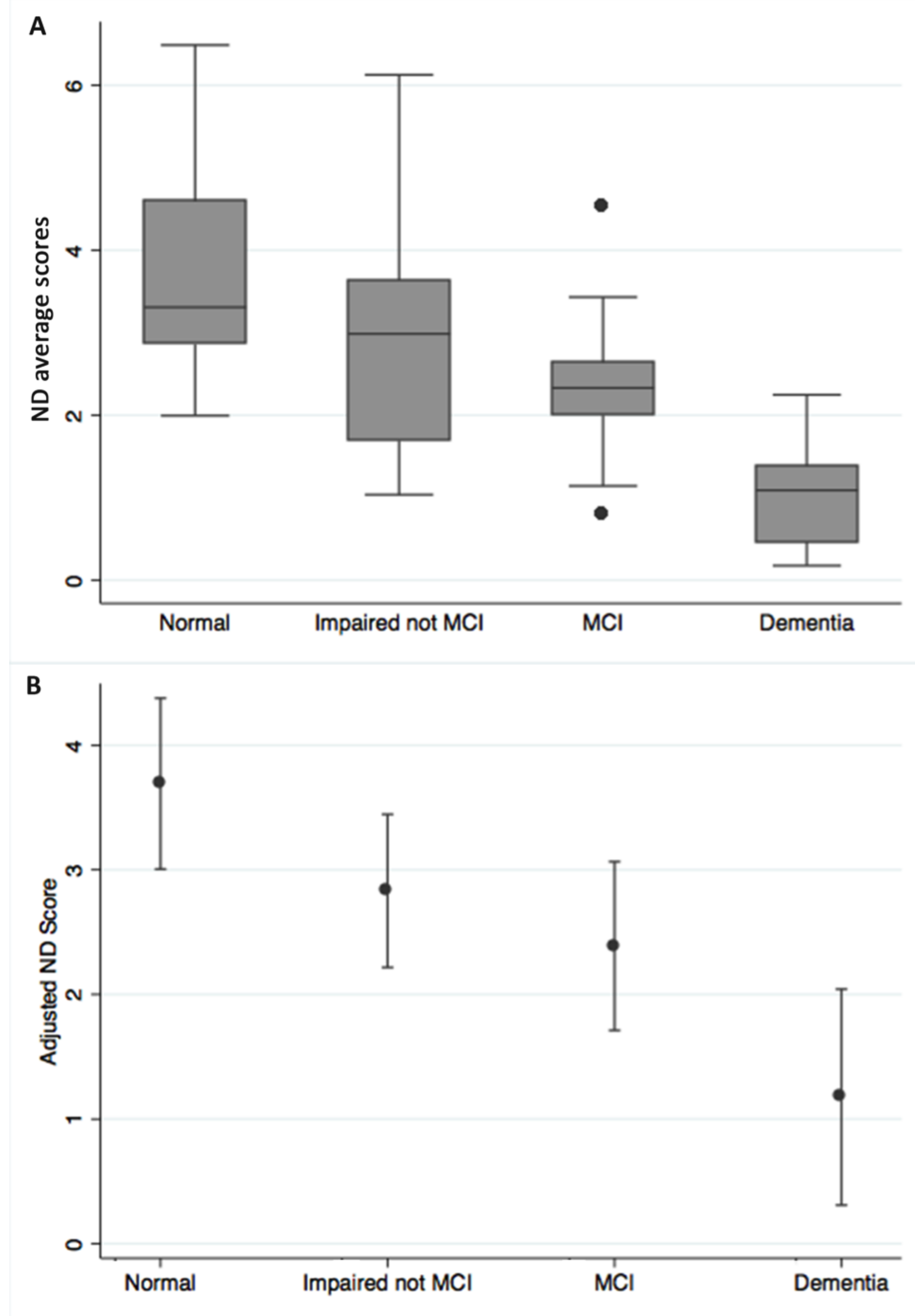


Figure 1. A) Box plot (top) depicting average ND scores for each participant by group. B) Adjusted ND scores (bottom) with 95% confidence intervals depicting predicted scores by group from a mixed model that controlled for covariates age, education, and gender.

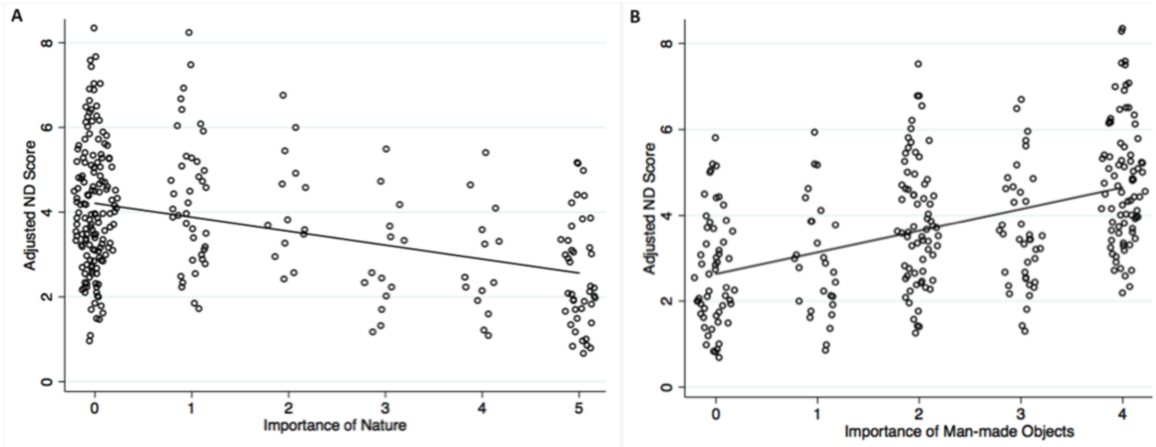


Figure 2. Narrative Description (ND) scores, adjusted for clip difficulty for the cognitively normal group depicting the importance of nature (left) or objects (right) was rated for understanding the clip (see section 2.4). Feature ratings ranged from 0 to 5. In this set of 20 clips, ranking 5 was not used for faces or man-made objects. To facilitate viewing the data, random offsets were applied in the importance dimension (abscissa) in each panel.

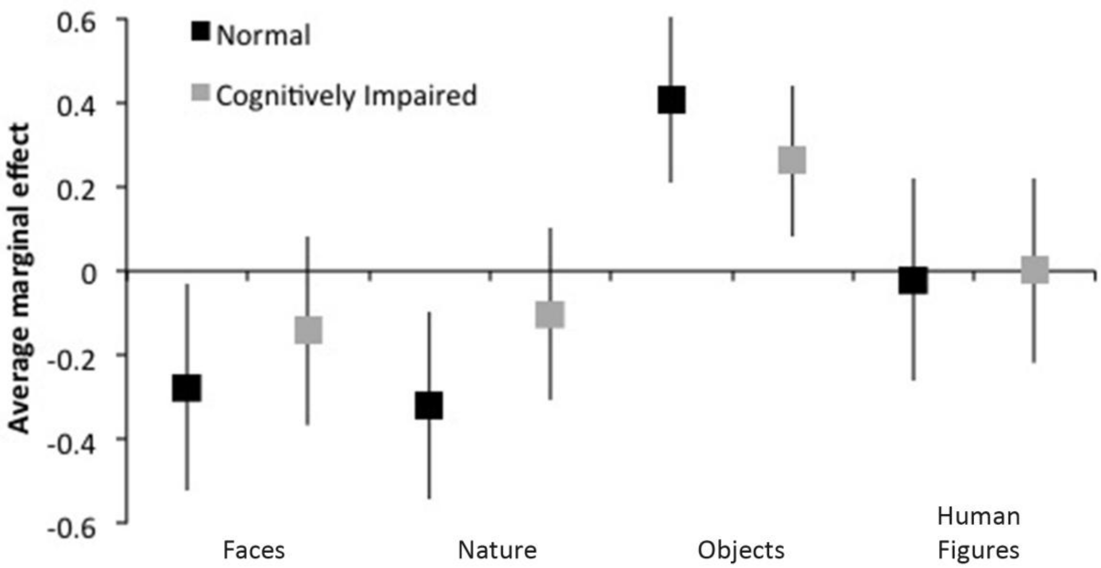


Figure 3. The effects of video characteristics on Narrative Description (ND) scores varied between the two groups, with the effect of video characteristic being less in the cognitively impaired group. The video characteristics were ratings (from 0 to 5) of the importance of each category for understanding of the video clip. The average marginal effects ("slopes") were found using a linear mixed model. Error bars are 95% confidence intervals. Significant effects are reported in the text.

Table 1. Cognitive status categories determined from diagnosis.

| Cognitive Status | Diagnosis |
|---|---|
| Normal (n=14) | Normal |
| Impaired-Not-MCI (n=19) | Impaired not MCI, MCIAPLUS, MCINON1 |
| MCI (n=15) | MCI, Amnestic MCI, MCINON2 |
| Dementia (n=8) | Dementia, Amnestic Multidomain Dementia, DEM-AMNDEM |

Mild Cognitive Impairment (MCI); Amnestic MCI plus one or more domains (MCIAPLUS); Non-amnestic MCI (MCINON); Amnestic Dementia (DEM-AMNDEM)

Table 2. Participant Demographics and Characteristics (mean and SD). Significance (Sig.) values represent Spearman correlations between the test and the levels of cognitive impairment, save for the variable gender for which the Chi-Square test was used.

| | **Cognitively Normal (n=14)** | **Impaired – Not MCI (n=19)** | **MCI (n=15)** | **Dementia (n=8)** | **Sig.** |
|---|---|---|---|---|---|
| Age (years) | 80(4.8) | 79(6.5) | 85(6.7) | 85(9.1) | p=0.006 |
| Gender (M:F) | 3:11 | 9:10 | 7:8 | 5:3 | p=0.74 |
| Education (years) | >16 years | >16 years | >16 years | 16 years | p=0.38 |
| MoCA | 27.5(1.4) | 25(3.5) | 23.3(3.3) | 18.6(4.0) | p<0.001 |
| CDR score | 0(0.1) | 0.4(0.2) | 0.4(0.3) | 1(0) | p<0.001 |
| CDR-SB | 0(0) | 1.1(0.9) | 2.1(0.9) | 4.75(0.7) | p<0.001 |
| Number Span Backwards | 5.1(1.5) | 5(1.6) | 4.5(1.1) | 4.4(0.7) | p=0.09 |
| MINT Total Score | 29.9(1.9) | 29.2(4.3) | 28.6(3.0) | 24.8(6.2) | p=0.01 |
| Category Fluency (animals + vegetables) | 35.6(7.6) | 33.3(10.6) | 26.4(7.3) | 17(9.0) | p<0.001 |
| Letter Fluency (F + L) | 33(10.1) | 27.2(9.5) | 27.7(9.8) | 23.8(6.6) | p=0.02 |
| TMT A (time) | 30.1(7.2) | 31.8(12.4) | 39.7(13.7) | 48.3(18.5) | p=0.005 |
| TMT B (time) | 67.2(24.3) | 92.3(55.1) | 131.7 (53.8) | 165.8 (76.4) | p<0.001 |
| TMT B-A (time) | 37.1(19.6) | 60.4(46.5) | 92(50.7) | 123(58.5) | p<0.001 |
| Craft Story 21 - Immediate | 17.4(5.3) | 14.6(4.1) | 10.9(4.5) | 5.6(1.8) | p<0.001 |
| Craft Story 21 - Delayed | 16.5(5.6) | 12.8(4.4) | 8.9(5.3) | 1.8(3.0) | p<0.001 |
| Craft Story 21 - % Retention | -85.7 (179.1) | -161.1 (211.8) | -221.4 (444.1) | -360 (336.2) | p=0.18 |
| Benson Figure - Copy | 16.8(0.6) | 16.2(1.5) | 15.7(0.9) | 14.5(3.3) | p<0.001 |
| Benson Figure - Delay | 12.3(2.5) | 10.1(3.3) | 8.5(4.8) | 4(3.4) | p<0.001 |
| Benson Figure - % Retention | -450(262) | -610(296) | -726.7 (531.1) | -1140 (472.2) | p=0.003 |
| Selective Reminding Test - Free Recall | 29.4(9.5) | 24.4(9.3) | 17.3(12.9) | 4.1(7.1) | p<0.001 |
| ND Score (avg) | 3.7 (1.3) | 2.9 (1.4) | 2.3 (1.0) | 1.0 (0.7) | p<0.001 |
| ND Score Range | 2.0 – 6.5 | 1.0 – 6.1 | 0.8 – 4.5 | 0.2 – 2.2 | N/A |

Mild Cognitive Impairment (MCI); Significance (Sig); male (M); female (F); Montreal Cognitive Assessment (MoCA); Cognitive Dementia Rating (CDR); Multilingual Naming Test (MINT): Trail Making Test (TMT); Narrative Description (ND).

Table 3. Spearman pairwise correlations between ND average scores and neuropsychological measures. Correlation coefficients are depicted with 95% confidence intervals. ** indicates significance of p<0.001; * indicates significance of p<0.01.

| | **ND Correlation Coefficients** | **95% Confidence Intervals** |
|---|---|---|
| MoCA | 0.71** | 0.61 to 0.82 |
| CDR-SB | -0.45** | -0.68 to -0.22 |
| MINT | 0.52** | 0.33 to 0.72 |
| Category Fluency | 0.63** | 0.46 to 0.80 |
| Letter Fluency | 0.52** | 0.33 to 0.71 |
| Number Span Backward | 0.42* | 0.16 to 0.68 |
| TMT A | -0.35 | -0.58 to -0.12 |
| TMT B | -0.56** | -0.76 to -0.36 |
| TMT B-A | -0.57** | -0.78 to -0.36 |
| Craft Story 21 Immediate | 0.49** | 0.29 to 0.69 |
| Craft Story 21 Delayed | 0.47** | 0.26 to 0.68 |
| Craft Story 21 - Percent Retention | 0.27 | 0.01 to 0.53 |
| Benson Figure Copy | 0.25 | 0.00 to 0.51 |
| Benson Figure Delay | 0.50** | 0.28 to 0.72 |
| Benson Figure - Percent Retention | 0.41* | 0.16 to 0.66 |
| Selective Reminding Test | 0.39 | 0.12 to 0.62 |

Narrative Description (ND); Montreal Cognitive Assessment (MoCA); Cognitive Dementia Rating Sum of Boxes (CDR-SB); Multilingual Naming Test (MINT): Trail Making Test (TMT); Selective Reminding Test – Free Recall (Free).

Table 4. Final stepwise ordinal logistical model predicting cognitive status from ND scores, CDR-SB, neuropsychological subtests, and demographic variables.

| Block | LL | LR | df | Pr >LR | AIC | BIC |
|---|---|---|---|---|---|---|
| 1 | -26.49 | 6.69 | 9 | 0.6689 | 80.98 | 107.47 |
| 2 | -29.84 | 66.52 | 2 | 0.0000 | 69.68 | 79.13 |

**Block 1**: Model began with ND average scores, CDR-SB, MINT, Benson (percent retention), Category Fluency (animals + vegetables), Craft Story (delayed recall), Trail Making Test (A), Trail Making Test (B-A), age, gender, and education. Trail Making Test (A) was removed first, followed by Category Fluency, gender, education, MINT, age, Benson (percent retention), Trail Making Test (B-A), and finally Craft Story (delayed recall). **Block 2**: The final model that ended with ND average scores and CDR-SB. Log-likelihood (LL); Likelihood-ratio test (LR); degrees of freedom (df); Probability (Pr); Akaike information criterion (AIC); Bayesian information criterion (BIC).

Table 5. Means and standard deviations of participant ratings on three questions pertaining to subjective experience of each video clip. Ratings were based on a scale from 0 (no interest / didn't like) to 5 (a lot of interest / liked a lot) for the first two questions. The third question, "seen the movie", required a binary response and is reported as the proportion seen.

| | **Normal** | **Impaired** | **MCI** | **Dementia** | **Sig.** |
|---|---|---|---|---|---|
| Interest in Movie | 2.8 (1.3) | 2.6 (1.3) | 2.1 (1.2) | 2.3 (1.3) | 0.015 |
| Liked the movie | 3.6 (0.9) | 3.5 (0.9) | 3.4 (0.8) | 3.4 (1.0) | 0.17 |
| Seen the movie | 27/274 (10%) | 30/356 (8%) | 10/292 (3%) | 16/160 (10%) | 0.02 |

Mild Cognitive Impairment (MCI); Significance (Sig)

# 5. References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008

Alzheimer's Association. (2016). 2016 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 12*(4), 459-509.

Aramaki, E., Shikata, S., Miyabe, M., & Kinoshita, A. (2016). Vocabulary size in speech may be an early indicator of cognitive impairment. *PLoS One, 11*(5), e0155195. doi:10.1371/journal.pone.0155195

Au, R., Piers, R. J., & Devine, S. (2017). How technology is reshaping cognitive assessment: Lessons from the Framingham Heart Study. *Neuropsychology, 31*(8), 846-861. doi:10.1037/neu0000411

Bayles, K. A., Tomoeda, C. K., Kaszniak, A. W., Stern, L. Z., & Eagans, K. K. (1985). Verbal perseveration of dementia patients. *Brain and Language, 25*(1), 102-116.

Besser, L., Kukull, W., Knopman, D. S., Chui, H., Galasko, D., Weintraub, S., . . . Clinical Core leaders of the National Institute on Aging-funded, U. S. A. s. D. C. (2018). Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. *Alzheimer's Disease & Associated Disorders, 32*(4), 351-358. doi:10.1097/WAD.0000000000000279

Braak, H., & Braak, E. (1997). Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiology of Aging, 18*(4), 351-357.

Brawley, E. C. (2007). Designing successful gardens and outdoor spaces for individuals with Alzheimer's Disease. *Journal of Housing For the Elderly, 21*(3-4), 265-283.

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., . . . Lyoo, C. H. (2016). In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. *Annals of Neurology, 80*(2), 247-258. doi:10.1002/ana.24711

Costela, F. M., Saunders, D. R., Kajtezovic, S., Rose, D. J., & Woods, R. L. (2018). Measuring the difficulty watching video with hemianopia and an initial test of a rehabilitation approach. *Translational Vision Science and Technology, 7*(4), 13.

Costela, F. M., Saunders, D. R., Rose, D. J., Katjezovic, S., Reeves, S. M., & Woods, R. L. (2019). People with central vision loss have difficulty watching videos. *Investigative Ophthalmology Visual Science, 60*(1), 358-364. doi:10.1167/iovs.18-25540

Costela, F. M., & Woods, R. L. (2018). When watching video, many saccades are curved and deviate from a velocity profile model. *Frontiers in Neuroscience, 12*, 960. doi:10.3389/fnins.2018.00960

Costela, F. M., & Woods, R. L. (2019). A free database of eye movements watching "Hollywood" videoclips. *Data in Brief, 25*, 103991. doi:10.1016/j.dib.2019.103991

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences, 26*(4), 243-254. doi:10.1007/s10072-005-0467-9

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's Disease in narrative speech. *Journal of Alzheimer's Disease, 49*(2), 407-422. doi:10.3233/JAD-150520

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders.* Philadelphia, PA: Lea & Febiger.

Graham, D. J., Stockinger, S., & Leder, H. (2013). An Island of Stability: Art images and natural scenes - but not natural faces - show consistent esthetic response in Alzheimer's-related dementia. *Frontiers in Psychology, 4*, 107. doi:10.3389/fpsyg.2013.00107

Hernandez-Dominguez, L., Ratte, S., Sierra-Martinez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 10*, 260-268. doi:10.1016/j.dadm.2018.02.004

Hodges, J. R., Salmon, D. P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: failure of access or degraded knowledge? *Neuropsychologia, 30*(4), 301-314.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry, 140*, 566-572. doi:10.1192/bjp.140.6.566

Kave, G., & Goral, M. (2016). Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 38*(9), 958-966. doi:10.1080/13803395.2016.1179266

Kortum, P., & Sullivan, M. (2010). The effect of content desirability on subjective video quality ratings. *Human Factors, 52*(1), 105-118. doi:10.1177/0018720810366020

Laatu, S., Revonsuo, A., Jaykka, H., Portin, R., & Rinne, J. O. (2003). Visual object recognition in early Alzheimer's disease: deficits in semantic processing. *Acta Neurologica Scandinavica, 108*(2), 82-89.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 263-269. doi:10.1016/j.jalz.2011.03.005

Morris, J. C., Weintraub, S., Chui, H. C., Cummings, J., Decarli, C., Ferris, S., . . . Kukull, W. A. (2006). The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Disease and Associated Disorders, 20*(4), 210-216. doi:10.1097/01.wad.0000213865.09806.92

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology, 40*(9), 917-939. doi:10.1080/13803395.2018.1446513

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: the mnemonic value of animacy. *Psychological Science, 24*(10), 2099-2105. doi:10.1177/0956797613480803

Okereke, O. I., Copeland, M., Hyman, B. T., Wanggaard, T., Albert, M. S., & Blacker, D. (2011). The Structured Interview & Scoring Tool-Massachusetts Alzheimer's Disease Research Center (SIST-M): development, reliability, and cross-sectional validation of a brief structured clinical dementia rating interview. *Archives of Neurology, 68*(3), 343-350. doi:10.1001/archneurol.2010.375

Okereke, O. I., Pantoja-Galicia, N., Copeland, M., Hyman, B. T., Wanggaard, T., Albert, M. S., . . . Blacker, D. (2012). The SIST-M: predictive validity of a brief structured clinical dementia rating interview. *Alzheimer Disease and Associated Disorders, 26*(3), 225-231. doi:10.1097/WAD.0b013e318231cd30

Ossenkoppele, R., Schonhaut, D. R., Scholl, M., Lockhart, S. N., Ayakta, N., Baker, S. L., . . . Rabinovici, G. D. (2016). Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain, 139*(Pt 5), 1551-1567. doi:10.1093/brain/aww027

Pozueco, L., Alvarez, A., Garcia, X., Garcia, R., Melendi, D., & Diaz, G. (2017). Subjective video quality evaluation of different content types under different impairments. *New Review of Hypermedia and Multimedia, 23*(1), 1-28.

Rule, N. O., Freeman, J. B., & Ambady, N. (2013). Culture in social neuroscience: a review. *Social Neuroscience, 8*(1), 3-10. doi:10.1080/17470919.2012.695293

Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012). Abnormalities of connected speech in semantic dementia versus Alzheimer's disease. *Aphasiology, 26*(6), 847-866.

Saunders, D. R., Bex, P. J., Rose, D. J., & Woods, R. L. (2014). Measuring information acquisition from sensory input using automated scoring of natural-language descriptions. *Public Library of Science ONE, 9*(4), e93251.

Saunders, D. R., Bex, P. J., & Woods, R. L. (2013). Crowdsourcing a normative natural language data set: A comparison of Mechanical Turk and in-lab data collection. *Journal of Medical Internet Research, 15*(5), e100. doi:10.2196/jmir.2620

Slooter, A. J., Cruts, M., Kalmijn, S., Hofman, A., Breteler, M. M., Van Broeckhoven, C., & van Duijn, C. M. (1998). Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Archives of Neurology, 55*(7), 964-968.

Sullivan, M., Pratt, J., & Kortum, P. (2008). *Practical issues in subjective video quality evaluation: Human factors vs. psychophysical image quality evaluation. .* Paper presented at the Paper presented at the ACM International Conference Proceeding Series. Proceeding of the lst International Conference on Designing Interactive User Experiences for TV and Video, New York.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Society and Nations.* . New York, NY: Doubleday.

Thal, D. R., Rub, U., Schultz, C., Sassin, I., Ghebremedhin, E., Del Tredici, K., . . . Braak, H. (2000). Sequence of Abeta-protein deposition in the human medial temporal lobe. *Journal of Neuropathology & Experimental Neurology, 59*(8), 733-748.

Tippett, L. J., Blackwood, K., & Farah, M. J. (2003). Visual object and face processing in mild-to-moderate Alzheimer's disease: from segmentation to imagination. *Neuropsychologia, 41*(4), 453-468.

Tomoeda, C. K., Bayles, K. A., Trosset, M. W., Azuma, T., & McGeagh, A. (1996). Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. *Alzheimer Disease and Associated Disorders, 10*(4), 204-215.

Verma, M., & Howard, R. J. (2012). Semantic memory and language dysfunction in early Alzheimer's disease: a review. *International Journal of Geriatric Psychiatry, 27*(12), 1209-1217. doi:10.1002/gps.3766

Weintraub, S., Besser, L., Dodge, H. H., Teylan, M., Ferris, S., Goldstein, F. C., . . . Morris, J. C. (2018). Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). *Alzheimer Disease and Associated Disorders, 32*(1), 10-17. doi:10.1097/WAD.0000000000000223

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine, 2*(4), a006171. doi:10.1101/cshperspect.a006171