

RESEARCH ARTICLE

# The Cognitive Online Self-Test Amsterdam (COST-A): Establishing norm scores in a community-dwelling population

Leonie N.C. Visser<sup>1,2</sup> | Mark A. Dubbelman<sup>1</sup> | Merike Verrijp<sup>1</sup> | Lisa Wanders<sup>3,4</sup> |  
Sophie Pelt<sup>1</sup> | Marissa D. Zwan<sup>1</sup> | Dick H.J. Thijssen<sup>3</sup> | Hans Wouters<sup>5</sup> |  
Sietske A.M. Sikkes<sup>1,6</sup> | Hein P.J. van Hout<sup>7</sup> | Wiesje M. van der Flier<sup>1,8</sup>

<sup>1</sup> Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands

<sup>2</sup> Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

<sup>3</sup> Radboud Institute for Health Sciences, Department of Physiology, Radboud University Medical Center, Nijmegen, The Netherlands

<sup>4</sup> Top Institute Food and Nutrition, Wageningen, The Netherlands

<sup>5</sup> General Practitioners Research Institute, Groningen, The Netherlands

<sup>6</sup> Faculty of Behavioural and Movement Sciences, Clinical Developmental Psychology & Clinical Neuropsychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>7</sup> Department of General Practice and Medicine for Older Persons, Amsterdam Institute for Public Health Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>8</sup> Department of Epidemiology and Biostatistics, Amsterdam UMC, Amsterdam, The Netherlands

## Correspondence

Mark A. Dubbelman, Alzheimer Center Amsterdam, Department of Neurology, Amsterdam UMC, VUmc, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands.  
Email: [m.dubbelman@amsterdamumc.nl](mailto:m.dubbelman@amsterdamumc.nl)

Leonie N.C. Visser and Mark A. Dubbelman contributed equally to this work.

Hein P.J. van Hout and Wiesje M. van der Flier contributed equally to this work.

## Funding information

Stichting Stoffels-Hornstra, ZonMw, Grant/Award Number: #73305095007

## Abstract

**Background:** Heightened public awareness about Alzheimer's disease and dementia increases the need for at-home cognitive self-testing. We offered Cognitive Online Self-Test Amsterdam (COST-A) to independent groups of cognitively normal adults and investigated the robustness of a norm-score formula and cutoff.

**Methods:** Three thousand eighty-eight participants (mean age  $\pm$  standard deviation =  $61 \pm 12$  years, 70% female) completed COST-A and evaluated it. Demographically adjusted norm scores were the difference between expected COST-A scores, based on age, gender, and education, and actual scores. We applied the resulting norm-score formula to two independent cohorts.

**Results:** Participants evaluated COST-A to be of adequate difficulty and duration. Our norm-score formula was shown to be robust:  $\approx 8\%$  of participants in two cognitively normal cohorts had abnormal scores. A cutoff of  $-1.5$  standard deviations proved optimal for distinguishing normal from impaired cognition.

**Conclusion:** With robust norm scores, COST-A is a promising new tool for research and clinical practice, providing low cost and minimally invasive remote assessment of cognitive functioning.

## KEYWORDS

Alzheimer's disease, cognition, normative data, remote assessment, screener, self-testing

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association

## 1 | INTRODUCTION

Early detection of Alzheimer's disease (AD) is essential for timely and adequate patient management, as well as for maximizing the benefit of potential disease-modifying treatment and other preventive efforts.<sup>1</sup> Likewise, easily accessible assessment of cognition has become increasingly relevant, with initiatives to raise public awareness about AD and other forms of dementia having heightened the vigilance for cognitive decline among the general public. This has left many concerned, yet non-demented, individuals in need of reassurance, information, and advice.<sup>2-4</sup> Furthermore, many people, and elderly in particular, experience physical, social, and psychological barriers hindering them from visiting their general practitioner or a memory clinic.<sup>5,6</sup> Consequently, many cases of imminent dementia remain undetected until later disease stages.<sup>7,8</sup>

Online testing could help to overcome these barriers, not in the least because of its flexibility in time and location, that is, the potential of self-administration at home.<sup>9</sup> The relevance of self-testing in the safety of one's own home is emphasized by the coronavirus disease 2019 (COVID-19) pandemic, which has impacted the availability and provision of face-to-face care to memory clinic patients.<sup>10,11</sup> Cognitive online self-testing removes the psychological obstacles associated with visiting a doctor or clinic and is low-cost. Nevertheless, despite its potential value, there are only a few online self-tests of cognitive functioning available to date (one example: the Amsterdam Cognition Scan<sup>12</sup>).

We previously developed an online self-administered test of cognitive functioning: the Cognitive Online Self-Test Amsterdam (COST-A), based on the Telephonic Remote Evaluation of Neuropsychological Deficits (TREND).<sup>13</sup> The content of the TREND was expanded by including visual and visuospatial tasks, resulting in a battery of 10 tasks measuring various cognitive abilities.<sup>14</sup> In a first validation study,<sup>14</sup> we tested the convergent validity and diagnostic accuracy of COST-A for mild cognitive impairment (MCI) and dementia in a memory clinic setting and found strong associations with paper-and-pencil cognitive tests, namely, the Mini-Mental State Examination (MMSE) (correlation coefficient  $r = .64$ ) and neuropsychological tests for memory ( $r = .71$ ), executive functioning ( $r = .57$ ), and attention ( $r = .5$ ).<sup>14</sup> COST-A also had adequate diagnostic accuracy in distinguishing MCI and dementia from cognitively normal individuals.<sup>14</sup>

The ultimate goal of COST-A is to enable community-dwelling older adults to self-assess their cognitive functioning and to promote early detection of cognitive decline in the context of AD and dementia. To better interpret COST-A performance of community-dwelling adults, representative normative data from online, self-administered, home assessment are required. Such data should be adjusted for age, education, and gender, as these factors are often found to be associated with cognitive performance in aging individuals.<sup>15</sup>

In this study, we collected COST-A data in a large, community-dwelling sample of adults to establish a formula for calculating demographically adjusted, standardized norm scores, and collect user experiences. In addition, we applied these norm scores in independent

### Research in Context

1. **Systematic review:** Online cognitive self-testing is a fairly novel practice. We reviewed results from a PubMed search and found that little is currently known about the usability and quality of online cognitive self-tests, and that normative data are mostly lacking. We previously developed an online self-administered test of cognitive functioning: the Cognitive Online Self-Test Amsterdam (COST-A).
2. **Interpretation:** Our current study provides demographically adjusted normative data for COST-A, a new cognitive self-assessment tool for research and clinical practice. This allows for substantive interpretation of test results and comparison to cognitively normal individuals.
3. **Future directions:** Improving communication about test results is an important avenue for future research. Furthermore, the predictive quality and usability of COST-A for the measurement of disease progression should be investigated in longitudinal studies.

samples of cognitively normal adults and patients diagnosed with MCI and mild dementia, to investigate their robustness.

## 2 | METHODS

### 2.1 | Study design

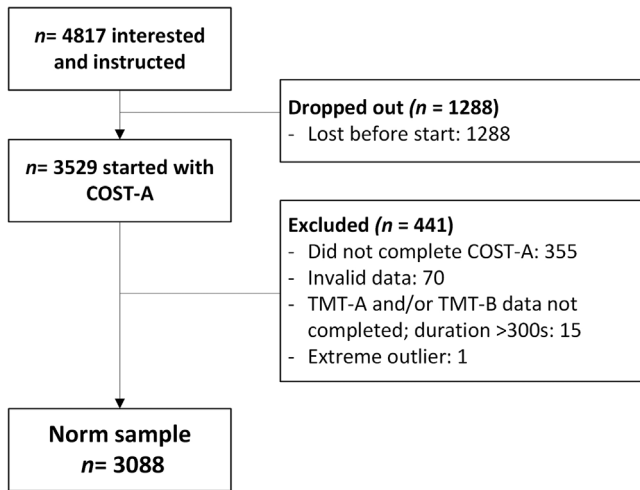
We recruited adults via the Dutch Brain Research Registry, an online platform for people interested in participating in brain-related research.<sup>16</sup> Data collection ran from August 2018 to December 2018 and was approved by the medical ethical committee of the VU University Medical Center. Participants provided consent via the Dutch Brain Research Registry.

Registrants who indicated their interest in participation were invited by e-mail to an online portal that hosted the cognitive test. Participants who did not respond were reminded two and four weeks after their first invitation, as well as approximately two weeks before the end of data collection.

### 2.2 | Participants and procedures

We invited individuals 18 years or older who had a good command of the Dutch language. Individuals who reported a dementia-related diagnosis (ie, dementia or MCI) were excluded.

The inclusion flow chart of the study is shown in Figure 1. At the time of recruitment, 11,060 registrants of the Dutch Brain Research Registry were eligible and invited to participate in this study by email.



**FIGURE 1** Flow chart of participant inclusion. Note: Of all participants who started with the COST-A ( $N = 3529$ ), 441 (12%) did not successfully complete all subtasks or were excluded based on invalid data, eg, because of technical issues. We used a  $t$  test for age and chi-square test for gender and education level to analyze differences in demographics between participants included in the norm sample ( $N = 3088$ ) and those who started but were excluded ( $N = 441$ ). The sample of excluded participants did not differ from the norm sample with respect to the distributions of gender (70% female,  $P = .183$ ) or education level (65% high,  $P = .133$ ), but excluded participants were significantly older ( $P < .001$ ), with a mean age of 64 years ( $SD = 13$ ). Abbreviations: COST-A, Cognitive Online Self-Test Amsterdam; TMT, Trail-Making Test

Of these, 4817 (44%) declared their interest. Complete and valid data were available from 3088 (64%). We refer to these 3088 participants as our “norm sample.”

## 2.3 | Measures

### 2.3.1 | Sample characteristics

Participants self-reported their age, gender, and level of education. We dichotomized education level into low–medium (up to the equivalent of high school education) and high education (the equivalent of college education).

### 2.3.2 | Cognitive Online Self-Test Amsterdam (COST-A)

The Cognitive Online Self-Test Amsterdam (COST-A) was developed by Van Mierlo et al.<sup>14</sup> COST-A was designed to be completed on a desktop or laptop computer, was hosted online by Neurotask, and was accessible via a personalized link to their website. Prior to starting COST-A, participants were instructed to (1) find a place with quiet surroundings, (2) reserve at least 20 minutes to complete COST-A in one go, and (3) to not use any aids, such as a calendar, clock, pen, or paper. COST-A takes approximately 20 minutes to complete and includes 10 cognitive tasks:

orientation (score 0–5), digit-sequence learning (score 0–3), immediate word recall (score 0–10), two trail-making tasks (A: connecting numbered dots, and B: alternately connecting lettered and numbered dots; score 0–300, indicating seconds to complete the task), delayed word recall (score 0–10), delayed word recognition (score 0–20), word pairs immediate recall (score 0–20), word pairs recognition (score 0–10), and semantic comprehension (score 0–6).

Raw scores on all subtasks were converted into standardized scores (Z-scores: mean 0,  $SD = 1$ ). Trail-making test scores were reverse-scored, so that higher scores represented better performance. The subtask Z-scores were then averaged into a composite score representing overall cognitive functioning. The resulting mean composite score is 0, with scores above 0 indicating better cognitive performance, and scores below 0 indicating poorer cognitive performance. Actual duration, that is, the time it took to complete all subtasks of the COST-A, is registered automatically.

### 2.3.3 | Additional measures

A single yes-or-no question assessed whether the participant experienced memory complaints. Depressive symptoms of participants were assessed with the five-item short form of the Geriatric Depression Scale (GDS-5),<sup>17</sup> which was administered online after completion of COST-A. Scores ranged from 0 to 5, with higher scores indicating more depressive symptoms. GDS-5 total scores were dichotomized into absence (total score of 0) or presence of depressive symptoms (total score of 1 or higher).

In addition, participants answered questions about their experiences during the completion of COST-A, including items evaluating difficulty, enjoyment, and duration of COST-A, and clarity of test instructions. Furthermore, four items assessed encountered completion issues: presence or absence of problems with vision, input (regarding the use of keyboard and mouse), interruptions, and whether participants perceived their surroundings as quiet, neutral, or noisy.

## 2.4 | Analyses

Analyses were performed in IBM SPSS version 26<sup>18</sup> and R version 4.0.3.<sup>19</sup>

### 2.4.1 | Demographically adjusted norm score formula

Before calculating a norm score formula, outliers on the composite Z-score ( $\leq -3$  or  $\geq 3$ ) were identified and excluded from data analysis to limit the influence of extreme scores. In regression analyses, we then regressed the composite cognition score on the predictive demographic variables age, gender (0 = male, 1 = female), and education (0 = low/medium, 1 = high), first in separate models, and then in a multiple regression model combining all demographic variables that were

significant predictors. The calculation of the norm scores is shown step-by-step below:

1. Compute the actual COST-A composite score by averaging subtask Z-scores using the means and SDs of the entire norm sample.
2. Use the coefficients from the multiple regression model containing significant demographic variables to compute an expected COST-A composite score.
3. Subtract the expected composite score from the actual composite score to obtain a difference score.
4. Standardize the difference score by dividing it by the residual standard error from the multiple regression model to obtain a demographically adjusted norm score.

We predetermined that a demographically adjusted norm score was abnormal if it was  $\leq -1.5$ . Distributions of the scores obtained in the four steps above are shown in the [Supplementary Material](#).

### 2.4.2 | Self-reported problems and user experiences

Descriptives were obtained for actual test duration, self-reported memory problems, depressive symptoms, encountered problems, and evaluation items. Using linear regression, we assessed the predictive value of self-reported memory problems and depressive symptoms on demographically adjusted COST-A norm scores. Next, we tested the influence of self-reported problems during COST-A completion on demographically adjusted norm scores.

### 2.4.3 | External validation of the COST-A norm score formula

We used two other data sets to validate our demographically adjusted norm scores and cutoff. First, of 2777 individuals who participated in the Nijmegen Exercise Study,<sup>20</sup> we selected 2440 individuals who did not have a dementia-related diagnosis and who successfully completed COST-A between June and November 2018. These individuals comprise our “validation sample,” as we consider them to represent a group of cognitively normal adults similar to our norm sample. Second, we selected 67 patients diagnosed with MCI or dementia from a previous sample recruited at the Alzheimer Center Amsterdam between February and October 2015.<sup>14</sup> This group comprised our “clinical sample.” In both these samples, participants completed COST-A at home without direct supervision of study personnel.

In both samples we computed the demographically adjusted norm scores as described earlier, using the formula based on our norm sample. We applied the  $-1.5$  SD cutoff to calculate the percentage of individuals who had a score below the mean. Finally, we calculated the optimal cutoff for distinguishing cognitively normal participants from persons with MCI or dementia based on the highest Youden index, representing the optimal balance between specificity and sensitivity, to compare with our predetermined cutoff.

**TABLE 1** Non-transformed (raw) scores on COST-A subtasks for the norm sample

Task	Mean	Standard deviation	Minimum	Maximum
1. Orientation	4.71	0.51	2	5
2. Digit-sequence learning	2.73	0.59	0	3
3. Immediate word recall	6.23	1.90	0	10
4. Connecting numbered dots (TMT-A)	37.94	17.21	6.74	262.46
5. Letter-number alteration (TMT-B)	66.45	30.08	20.41	296.56
6. Free delayed word recall	5.28	2.20	0	10
7. Delayed word recognition	18.61	1.46	10	20
8. Word pairs immediate recall	12.06	4.24	0	20
9. Word pairs recognition	9.25	1.29	0	10
10. Semantic comprehension	5.79	0.50	2	6

Tasks 4 and 5 are scored as seconds needed to complete; all other tasks are scored as number of correct responses. Tasks 4 and 5 are reverse scored before being standardized for the composite score.

## 3 | RESULTS

### 3.1 | Sample characteristics

A total of 3088 individuals ( $60.6 \pm 12.1$  years old, range 18-96 years; 70.0% female; 68.2% highly educated) completed COST-A. Age distributions stratified by gender and education level can be found in the [Supplementary Material](#). It took the participants on average 17 minutes (SD = 6 minutes) to complete COST-A. Table 1 presents descriptive statistics of raw scores of the 10 COST-A subtasks for the norm sample.

### 3.2 | Demographically adjusted composite score formula

The regression models, presented in Table 2, provide regression-based normative data. The betas show that higher age, male gender, and low/medium education level significantly deteriorated composite cognition scores, both in separate linear regressions models, as well as in a multiple linear regression model including all demographic variables (all  $P < .001$ ).

Box 1 illustrates the four steps required to obtain the demographically adjusted norm scores. After computing the COST-A composite

**TABLE 2** Regression-based normative data

Model	Variable	Beta	95%CI	SD residual
<b>Simple regressions</b>				
	Age	-0.019	[-0.020, -0.017]	0.502
	Gender	0.215	[0.173, 0.256]	0.540
	Education	0.338	[0.298, 0.378]	0.526
<b>Multiple regression</b>				
	Constant	0.666	[0.563, 0.769]	0.479
	Age	-0.016	[-0.018, -0.015]	
	Gender	0.158	[0.120, 0.195]	
	Education	0.296	[0.260, 0.333]	

Age is entered into the models as age in years; gender as 0 = male, 1 = female; education as 0 = low/medium, 1 = high. Abbreviations: CI, confidence interval; SD, standard deviation.

score, the intercept and unstandardized betas derived from the multiple regression model provided the following formula for calculating the expected composite score:  $0.666 - 0.016 \times \text{Age (years)} + 0.158 \times \text{Gender (0 = male, 1 = female)} + 0.296 \times \text{Education (0 = low/medium, 1 = high)}$ . The difference between the expected and actual COST-A composite score is divided by the residual standard error from the multiple linear regression model (0.479) to obtain a demographically adjusted, standardized norm score. In our norm sample, 242 individuals (7.8%) had a norm score below the cutoff of -1.5 SD.

### 3.3 | Self-reported problems and user experiences

In the norm sample, self-reported memory problems (reported by 47.3%) and depressive symptoms ( $M = 0.6$ ,  $SD = 1.0$ , range = 0-5, 34.2% with score 1 or higher) were associated with cognitive functioning. Participants who reported memory problems performed worse ( $P < .001$ ) on COST-A (demographically adjusted norm score =  $-0.16 \pm 1.06$ ) than participants who did not report problems ( $0.15 \pm 0.92$ ). In addition, their scores were more often below the cutoff (11.3%) than participants who reported no memory problems (4.7%;  $\chi^2(1) = 45.36$ ,  $P < .001$ ). Patients who reported depressive symptoms had lower cognitive functioning ( $-0.06 \pm 0.61$ ) than those who did not ( $0.03 \pm 0.51$ ;  $P < .001$ ). The scores of participants who reported depressive symptoms more often fell below the cutoff (11.4%) than the scores of participants who did not (6.0%;  $\chi^2(1) = 27.46$ ,  $P < .001$ ).

Table 3 displays data regarding participants' experiences with COST-A. Overall, participants' evaluations of the test and test instructions were positive; only a few reported the test to be difficult (7%), dreadful (1%), or long (2%), or the instructions unclear (< 1%).

Small proportions of the norm sample reported to have experienced problems with vision (1%) or the mouse/keyboard (4%), or to have completed the test in a noisy environment (2%). In addition, a larger percentage of participants reported to have been interrupted (13%) at some point during the test (Table 3). Participants who reported any

**TABLE 3** Test experiences and conditions

Aspect	N* (%)
<b>Difficulty</b>	
Easy	814 (26.5)
Neutral	2054 (66.8)
Difficult	207 (6.7)
<b>Judgment</b>	
Fun	2510 (81.6)
Neutral	544 (17.7)
Dreadful	21 (0.7)
<b>Duration</b>	
Short	399 (13.0)
Neutral	2617 (85.1)
Long	59 (1.9)
<b>Instructions</b>	
Unclear	9 (0.3)
Neutral	89 (2.9)
Clear	2977 (96.8)
<b>Completion issues</b>	
Vision problems	24 (0.8)
Input problems (mouse or keyboard)	119 (3.9)
Interrupted	408 (13.3)
<b>Surroundings</b>	
Quiet	2145 (69.8)
Neutral	863 (28.1)
Noisy	67 (2.2)

\*Data available from  $N = 3075$ .

Of note, all eight evaluation items were completed by the norm sample only: participants who successfully completed all subtasks of COST-A.

of these difficulties had lower norm scores ( $-0.22 \pm 1.0$ ) than those who did not encounter any difficulties ( $0.05 \pm 0.99$ ;  $P < .001$ ). Participants who encountered problems were also more likely to score below the cutoff (10.7%) than participants who did not (7.2%;  $\chi^2(1) = 7.11$ ,  $P = .008$ ). Of note, when participants were interrupted (the most frequently reported problem), they performed worse on COST-A ( $-0.20 \pm 1.00$  vs  $0.03 \pm 1.00$ ,  $P < .001$ ;) but their scores were not more often below the cutoff (10.0%) than those who were not interrupted (7.5% below cutoff,  $\chi^2(1) = 2.94$ ,  $P = .086$ ).

### 3.4 | Validation of norm-score formula in independent samples

The validation sample comprised 2440 community-dwelling individuals (mean age  $58.7 \pm 12.7$  years, 43.1% female, 54.4% highly educated). Compared to our original norm sample, individuals included in the validation sample were younger, more likely to be male, and less likely to be highly educated (all  $P$ 's < .001).

**Box 1**

**Formula demographically adjusted composite score** =  $0.666 - 0.016 * \text{Age} + 0.158 * \text{Gender} + 0.296 * \text{Education}$

Gender: 0 = male, 1 = female; Education: 0 = low/medium, 1 = high

**Example:** A 63-year-old, highly educated woman has the following raw scores, corresponding Z-scores, and actual composite score:

Task	Raw score	Z-score <sup>a</sup>
1. Orientation	5	0.57
2. Digit span forward	3	0.46
3. Immediate word recall	5	-0.65
4. Connecting numbered dots (TMT-A) in sec	45	-0.41 <sup>b</sup>
5. Letter number alternation (TMT-B) in sec	75	-0.28 <sup>b</sup>
6. Delayed word recall (free recall)	5	-0.13
7. Delayed word recognition	18	-0.42
8. Word pairs immediate recall	11	-0.25
9. Word pairs recognition	8	-0.97
10. Semantic comprehension	5	-1.58
<b>Actual composite score (z-score average)</b>		<b>-0.37</b>

<sup>a</sup>Computed using this formula:  $(\text{Raw score} - M) / \text{SD}$ ; M and SD values provided in Table 2.

<sup>b</sup>Z-score inverted (multiplied by -1)

Expected composite score (following formula):  $0.666 - 0.016 * 63 + 0.158 * 1 + 0.296 * 1 = 0.11$

Difference score  $(-0.37 - 0.11)$ :  $-0.48$

Norm score (difference score/SD residual):  $-0.48 / 0.479 = -1.00$ . She scores above the  $-1.5\text{SD}$  cutoff. Since this is a standardized norm score (following a normal data distribution), a score of  $-1$  means that 16% of women similar to her in age and education level performed the same or worse.

**What if this woman is age 40, instead of 63?** Then her expected composite score becomes 0.48, the difference score  $-0.85$  and her norm score  $-1.77$ . This score is abnormal, considering our  $-1.5$  cutoff. Only 4% of women similar to her in age and education level performed the same or worse.

Using the demographically adjusted norm score formula obtained from the original norm sample, norm scores were calculated for the validation sample. One hundred eighty-six individuals (7.6%) from the validation sample had a demographically adjusted norm score of  $\leq -1.5$ .

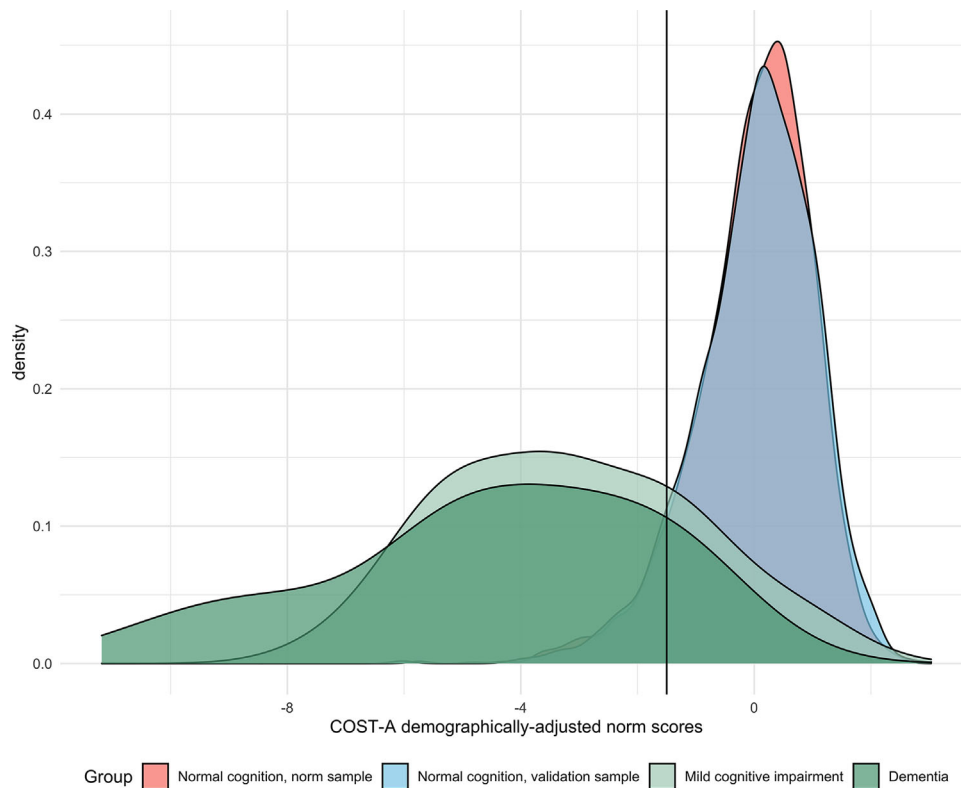
We subsequently included a sample of 67 memory clinic patients ( $65.0 \pm 7.3$  years of age, 71.6% female, 37.3% highly educated), with syndrome diagnoses of MCI ( $n = 28$ ) or dementia ( $n = 39$ ). Compared to the norm sample, both clinical groups performed worse on COST-A (MCI:  $-1.64$ , 95% confidence interval [CI]  $-1.86, -1.43$ ; dementia:  $-2.32$ , 95%CI  $-2.50, -2.15$ ). Based on our predetermined cutoff of  $-1.5\text{SD}$ , 21 MCI patients (75.0%) and 32 dementia patients (82.1%) had abnormal COST-A scores. With a maximized Youden index, the optimal cutoff on the demographically adjusted norm score for distinguishing cognitively normal individuals from cognitively abnormal

(MCI or dementia) was  $-1.5$  (accuracy = 91.6%, sensitivity = .81, specificity = .91).

Figure 2 shows the distribution of COST-A demographically adjusted norm scores in the separate cognitively healthy samples (the norm sample in red, the validation sample in blue), as well as in individuals with MCI and dementia (in shades of green)

## 4 | DISCUSSION

Based on assessments from 3088 cognitively normal Dutch adults, we established a regression-based formula to calculate demographically adjusted, standardized norm scores for the Cognitive Online Self-Test Amsterdam (or COST-A). We then used this formula to calculate demographically adjusted norm scores in 2440 community-dwelling adults.



**FIGURE 2** Demographically adjusted norm scores for COST-A, stratified by sample and diagnosis. The vertical line represents the predetermined cutoff of  $-1.5$  SD

In both samples, approximately 8% of participants scored below the cutoff, illustrating the robustness of our norm formula. When applying the norm formula to a clinical sample, 75% of MCI patients and 82% of dementia patients had abnormal demographically adjusted norm scores. The cutoff was also confirmed to distinguish cognitively normal from cognitively abnormal individuals (MCI or dementia) with high accuracy. As such, COST-A may serve as a screening tool to identify those individuals who require formal assessment by a health care professional and/or for research studies.

Most individuals who completed COST-A evaluated the test positively. Test completion took less than 20 minutes, which the large majority of participants judged to be acceptable. Of note, approximately 10% of cognitively normal participants who started COST-A did not successfully complete all tasks. This percentage was lower than in some smartphone-based assessments<sup>21,22</sup> and another online self-test,<sup>12</sup> but higher than in an online cognitive assessment tool that was framed as a game.<sup>23</sup> Participants who were unable to complete COST-A, and who were therefore excluded, were older than participants who did complete it. Still, our norm sample comprised sufficient numbers of older adults, up to 96 years of age, demonstrating feasibility of online self-administration of cognitive tests among the elderly.

As expected, and in accordance with previous findings,<sup>24–26</sup> older age and a lower education level significantly predicted lower cognitive functioning as measured with COST-A. In our norm sample, we

found that men performed worse than women. Previous studies have shown both better and worse cognitive performance in men,<sup>24</sup> and further exploration of these gender differences is thus an interesting direction for future research. Based on our findings, an adjustment for these demographic characteristics is warranted. The demographically adjusted norm scores thus represent the normalized difference between expected and actual performance on COST-A for a given individual. Following standard practice,<sup>24</sup> a norm score that was more than 1.5 SD below the mean was predetermined to represent abnormal performance. Our findings corroborated this cutoff, as it was also the optimal value to distinguish between cognitively normal individuals and those with MCI or dementia.

A few factors may impede the interpretation of COST-A results at the individual level. First, a proportion of participants reported to have been interrupted during the test, which can occur in the home environment. Although such interruption might negatively influence an individual's score, our results show that this does not necessarily result in an abnormal norm score. When considering any kind of encountered difficulties, including problems with the keyboard or vision, individuals' chances to score below the cutoff increased. This could, however, be a case of reverse causality, that is, we cannot discern whether these problems caused a lower test score, or whether those who have more cognitive problems encountered more problems. To resolve this issue of reverse causality, future research could invite participants who reported difficulties when completing COST-A during

self-administration at home for a face-to-face neuropsychological assessment in a more controlled environment, to allow comparison of performance between methods. Nevertheless, it is important to be mindful of the potential influence of self-reported encountered difficulties when interpreting an individual's result.

Second, we found that participants with more depressive symptoms perform worse and more often score below the cutoff, which is in accordance with previous research showing that individuals with depressive symptoms have poorer cognitive performance.<sup>27,28</sup> In addition, approximately half of our cognitively healthy participants reported memory complaints. This high number may, in part, be because we asked participants whether they experienced memory complaints immediately after completing COST-A. Still, participants who self-reported memory complaints, indeed had lower COST-A scores and were more likely to fall below the cutoff than participants who did not. For many, their subjective experience did thus match their objective performance. However, only one in five individuals who reported memory complaints actually scored below the cutoff. This emphasizes the importance of a short cognitive screening such as COST-A, as it can help to underscore or refute subjective cognitive complaints in comparison to others. Overall, at-home assessment seems feasible for many, including the elderly and participants with a low to medium level of education, if the aforementioned factors are considered when interpreting and disclosing an individual's score.

Online testing is now increasingly applied in both clinical practice and research.<sup>29,30</sup> It has numerous benefits over in-person testing, including the possibility for remote assessment. This is especially relevant in light of the COVID-19 pandemic, which has forced many studies to resort to alternative forms of cognitive testing. Even when government restrictions are eventually lifted, online testing will allow those who are unable to visit a memory clinic or research site to be tested remotely. In addition, it may facilitate follow-up of patients and participants over time by reducing the barrier of travel to the clinic or research site every time.

We do not consider an online self-test such as COST-A to be a replacement for other diagnostic tests such as those provided in a memory clinic. Nevertheless, with the evidence we present here, COST-A may play an important role in the diagnostic process for neurodegenerative diseases as a pre-screener. When an individual worries about their cognitive functioning, they could complete COST-A at home. When their COST-A score falls within the expected range, this might serve as reassurance. When it is below the cutoff, the individual could be encouraged to visit their general practitioner, who can refer them to a memory clinic if deemed appropriate. Furthermore, in the future, it is conceivable that COST-A may be combined with relatively low-invasive blood biomarker testing to perform pre-screening for clinical trials of disease-modifying drugs, as these are increasingly aimed at the earliest disease stages.<sup>31</sup>

This study had some limitations. We did not perform more extensive neuropsychological testing of the cognitively normal individuals, and as such it is possible that we may have included people with undiagnosed cognitive impairment in these samples. To reduce the influence of these

potential cases, we removed individuals with COST-A scores of more than 3 SD below the mean. In addition, because assessment was completely remote, we could not control the environment in which COST-A was completed, including test conditions and the device used. Although this may have introduced some noise into our data, we do believe this more accurately reflects the home environment in which the test may eventually be administered.

An important strength of this study was the inclusion of two large sets of cognitively normal individuals across a wide age range, providing an adequate representation of the Dutch adult population. This means that the norm-score formula can be applied broadly to both men and women of varying ages and education levels. Another strength was the inclusion of a clinical sample to obtain details about the discriminatory ability of COST-A. This sample included individuals with MCI and dementia, and their performance was markedly lower, proving that COST-A can be used to distinguish between cognitively normal and abnormal. Furthermore, we included an appraisal of participants user experience of COST-A, going beyond frequently reported measures such as compliance and completion rates.

This study provides many opportunities for future research. To further develop a test result report, including explanatory texts and visuals, we envision a qualitative study involving all stakeholders. This result report should counsel individuals in their interpretation of the results, and give advice on whether or not to visit a general practitioner based on the results. Then, the impact of implementing COST-A should be piloted and investigated in clinical practice, for example, the primary care setting, and the possibility of administering COST-A on mobile devices should be investigated. Finally, test-retest reliability as well as the predictive quality and usability of COST-A for the measurement of disease progression should be investigated in longitudinal studies.

In conclusion, we established a formula to obtain demographically adjusted norm scores for COST-A, a brief online self-test, and subsequently showed that these norms are robust across populations. As such, COST-A is a promising new tool for clinical and research practice, providing low-cost and minimally invasive remote assessment of cognitive functioning.

## ACKNOWLEDGMENTS

We would like to thank our participants for their involvement, valuable time, and effort. Research of Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc fonds.

Researchers who are interested in using COST-A should contact the authors for further details.

Funding for the data collection and research on which this study is based was provided by the Stichting Stoffels-Hornstra. Hersenonderzoek.nl was used to contact potential participants and is supported by a grant from ZonMW Memorabel (#73305095003).

WMF and LNCV are recipients of ABOARD, a public-private partnership receiving funding from ZonMw (#73305095007) and



Health~Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). LNCV is supported by a fellowship grant received from Alzheimer Nederland (WE.15-2019-05). LW is supported by a grant from the Netherlands Organization for Scientific Research (grant number: ALWTF.2016.032). SAMS is recipient of public-private funding from Health~Holland, Topsector Life Sciences & Health (PPP-allowance; LSHM20084-SGF, project DEFEAT-AD), and the National Institutes of Health, as well as license fees from Green Valley, VtV Therapeutics, Alzheon, Vivoryon, and Roche, and honoraria from Boehringer and Toyama. All funding is paid to her institution. WMF has received further funding from NWO, EU-FP7, EU-JPND, Alzheimer Nederland, CardioVascular Onderzoek Nederland, stichting Dioraphte, Gieskes-Strijbis fund, stichting Equilibrio, Pasman stichting, Biogen MA Inc, Boehringer Ingelheim, Life-MI, AVID, Roche BV, Fujifilm, and Combinostics. WMF holds the Pasman chair, has performed contract research for Biogen MA Inc. and Boehringer Ingelheim, is a consultant to Oxford Health Policy Forum CIC, Roche and Biogen MA Inc., and has been an invited speaker at Boehringer Ingelheim, Biogen MA Inc., Danone, Eisai, WebMD Neurology (Medscape). All funding is paid to her institution. HPJH received grants from European Committee, and Care Institute Netherlands, which were paid to his institution. The other authors have no relevant disclosures.

#### CONFLICT OF INTEREST

The authors report no conflict of interest.

#### REFERENCES

- Frisoni GB, Molinuevo JL, Altomare D, et al. Precision prevention of Alzheimer's and other dementias: anticipating future needs in the control of risk factors and implementation of disease-modifying therapies. *Alzheimers Dement*. 2020;16:1457-1468.
- Kunnean M, Pel-Littel R, Bouwman FH, et al. Patients' and caregivers' views on conversations and shared decision making in diagnostic testing for Alzheimer's disease: the ABIDE project. *Alzheimers Dement (N Y)*. 2017;3:314-322.
- Visser LNC, Kunnean M, Murugesu L, et al. Clinician-patient communication during the diagnostic workup: the ABIDE project. *Alzheimers Dement (Amst)*. 2019;11:520-528.
- Fruijtier AD, Visser LNC, van Maurik IS, et al. ABIDE Delphi study: topics to discuss in diagnostic consultations in memory clinics. *Alzheimers Res Ther*. 2019;11:77.
- Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Clin Neuropsychol*. 2012;26:177-196.
- Feenstra HE, Vermeulen IE, Murre JM, Schagen SB. Online cognition: factors facilitating reliable online neuropsychological test results. *Clin Neuropsychol*. 2017;31:59-84.
- Van den Dungen P, Van Marwijk HWM, Van der Horst HE, et al. The accuracy of family physicians' dementia diagnoses at different stages of dementia: a systematic review. *Int J Ger Psych*. 2012;27:342-354.
- Lang L, Clifford A, Wei L, et al. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *Bmj Open*. 2017;7:e011146.
- Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. *Clin Neuropsychol*. 2013;27:1328-1361.
- van Maurik IS, Bakker ED, van den Buuse S, et al. Psychosocial effects of corona measures on patients with dementia, mild cognitive impairment and subjective cognitive decline. *Front Psychiatry*. 2020;11:585686.
- Benaque A, Gurruchaga MJ, Abdelnour C, et al. Dementia care in times of COVID-19: experience at Fundacio ACE in Barcelona, Spain. *J Alzheimers Dis*. 2020;76:33-40.
- Feenstra HE, Vermeulen IE, Murre JM, Schagen SB. Online self-administered cognitive testing using the Amsterdam Cognition Scan: establishing psychometric properties and normative data. *J Med Internet Res*. 2018;20:e192.
- Mundt JC, Kinoshita LM, Hsu S, Yesavage JA, Greist JH. Telephonic Remote Evaluation of Neuropsychological Deficits (TREND): longitudinal monitoring of elderly community-dwelling volunteers using touch-tone telephones. *Alzheimer Dis Assoc Dis*. 2007;21:218-224.
- Van Mierlo LD, Wouters H, Sikkes SAM, et al. Screening for mild cognitive impairment and dementia with automated, anonymous online and telephone cognitive self-tests. *J Alzheimers Dis*. 2017;56:249-259.
- Van der Elst W, Molenberghs G, van Tetering M, Jolles J. Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *Clin Neuropsychol*. 2017;31:1173-1187.
- Zwan MD, van der Flier WM, Cleutjens S, et al. Dutch Brain Research Registry for study participant recruitment: design and first results. *Alzheimers Dement (N Y)*. 2021;7:e12132.
- Hoyl MT, Alessi CA, Harker JO, et al. Development and testing of a five-item version of the Geriatric Depression Scale. *J Am Geriatr Soc*. 1999;47:873-878.
- IBM Corp. *IBM SPSS Statistics for Windows*. 26 ed. Armonk, NY: IBM Corp.; 2019.
- R Core Team. R: A language and environment for statistical computing. 4.0.2 ed 2020.
- Maessen MF, Verbeek AL, Bakker EA, Thompson PD, Hopman MT, Eijsvogels TM. Lifelong exercise patterns and cardiovascular health. *Mayo Clin Proc*. 2016;91:745-754.
- Schweitzer P, Husky M, Allard M, et al. Feasibility and validity of mobile cognitive testing in the investigation of age-related cognitive decline. *Int J Methods Psychiatr Res*. 2017:26.
- Jongstra S, Wijsman LW, Cachucho R, Hoevenaar-Blom MP, Mooijaart SP, Richard E. Cognitive testing in people at increased risk of dementia using a smartphone app: the iVitality proof-of-principle study. *JMIR Mhealth Uhealth*. 2017;5:e68.
- Aalbers T, Baars MA, Olde Rikkert MG, Kessels RP. Puzzling with online games (BAM-COG): reliability, validity, and feasibility of an online self-monitor for cognitive performance in aging adults. *J Med Internet Res*. 2013;15:e270.
- Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological assessment*. 5th ed.. New York, NY: Oxford University Press; 2012.
- Borland E, Stomrud E, van Westen D, Hansson O, Palmqvist S. The age-related effect on cognitive performance in cognitively healthy elderly is mainly caused by underlying AD pathology or cerebrovascular lesions: implications for cutoffs regarding cognitive impairment. *Alzheimers Res Ther*. 2020;12:30.
- Bennett DA, Wilson RS, Schneider JA, et al. Education modifies the relation of AD pathology to level of cognitive function in older persons. *Neurology*. 2003;60:1909-1915.
- Steffens DC, Potter GG. Geriatric depression and cognitive impairment. *Psychol Med*. 2008;38:163-175.
- Leyhe T, Reynolds CF 3rd, Melcher T, et al. A common challenge in older adults: classification, overlap, and therapy of depression and dementia. *Alzheimers Dement*. 2017;13:59-71.

29. Mackin RS, Insel PS, Truran D, et al. Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: results from the Brain Health Registry. *Alzheimers Dement (Amst)*. 2018;10:573-582.
30. Perin S, Buckley RF, Pase MP, et al. Unsupervised assessment of cognition in the Healthy Brain Project: implications for web-based registries of individuals at risk for Alzheimer's disease. *Alzheimers Dement (N Y)*. 2020;6:e12043.
31. Crous-Bou M, Minguillon C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimers Res Ther*. 2017;9:71.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Visser LNC, Dubbelman MA, Verrijp M, et al. The Cognitive Online Self-Test Amsterdam (COST-A): Establishing norm scores in a community-dwelling population. *Alzheimer's Dement*. 2021;13:e12234.  
<https://doi.org/10.1002/dad2.12234>