



RESEARCH

Validity and diagnostic accuracy of the Auditory Consonant Trigram (ACT) Test in Turkish patients with mild cognitive impairment and Alzheimer's disease

İşitsel Üçlü Sessiz Harf Sıralaması (ACT) Testi'nin Türk hafif bilişsel bozukluk ve Alzheimer hastaları için geçerliği ve tanısal doğruluğu

Ata Cantürk Doğrul¹, Banu Cangöz Tavat¹, Esen Saka Topçuoğlu²

¹Hacettepe University, Ankara, Türkiye

²Private Neurology Clinic, Ankara, Türkiye

Abstract

Purpose: This study is a multi-method validity study aimed to evaluate the clinical usefulness of the Auditory Consonant Trigram (ACT) Test in distinguishing healthy controls, patients with Mild Cognitive Impairment (MCI), and patients with Alzheimer's disease (AD).

Materials and Methods: A total of 166 participants aged 50–86 years were included: 22 with AD, 20 with MCI, and 124 healthy controls. All participants first completed a battery of standardized neuropsychological tests, followed by the ACT Test.

Results: Significant group differences were found in ACT Test scores, indicating an impairment gradient across delay intervals (Control > MCI > AD). ACT Test total scores correlated moderately to strongly with general cognitive functions, memory, and executive functions. Logistic regression analysis found that ACT Test significantly distinguished patients with AD from controls (OR = 0.89, 95% CI [0.84–0.96]), but did not differentiate patients with MCI from controls (OR = 1.04, 95% CI [0.91–1.19]). In the extended model that included age and other cognitive variables, ACT Test was not an independent predictor. ROC analysis showed strong discrimination for AD (AUC = 0.89, 95% CI [0.83–0.94]) with a cut-off point of ≤ 39, sensitivity of 91% (95% CI [0.71–0.99]), and specificity of 82% (95% CI [0.74–0.89]). Adjusting for age further improved accuracy in the age-adjusted model.

Conclusion: The ACT Test provides robust discrimination of AD and shows meaningful convergent associations with multiple cognitive domains.

Keywords: Alzheimer's disease, mild cognitive impairment, working memory, neuropsychological tests, ROC curve

Öz

Amaç: Bu çalışma, İşitsel Üçlü Sessiz Harf Sıralaması (ACT) Testi'nin, sağlıklı bireyler ile Hafif Bilişsel Bozukluk (HBB) ve Alzheimer hastalarını (AH) ayırt etmedeki klinik yararlılığını değerlendirmeyi amaçlayan çok yöntemli bir geçerlik çalışmasıdır.

Gereç ve Yöntem: Yaşları 50-86 arasında değişen toplam 166 katılımcı (22 AH, 20 HBB ve 124 sağlıklı kontrol) çalışmaya dahil edilmiştir. Tüm katılımcılara standart bir nöropsikolojik test bataryasının ardından ACT Testi uygulanmıştır.

Bulgular: ACT Testi puanları tanı grupları arasında anlamlı farklılık göstermiş ve gecikme süreleri boyunca performansta belirgin bir bozulma eğilimi ortaya koymuştur (Kontrol > HBB > AH). ACT Testi toplam puanları, genel biliş ve bellek ve yürütücü işlevlerle orta-güçlü ilişkiler göstermiştir. Lojistik regresyon ACT Testi'nin AH hastaları ile kontrol grubunu ayırt ettiğini (OR = 0.89, 95% GA [0.84–0.96]) ancak HBB hastalarını ayırt etmediğini göstermiştir (OR = 1.04, 95% GA [0.91–1.19]). Yaş ve diğer bilişsel değişkenlerin dahil edildiği genişletilmiş modelde ACT Testi bağımsız bir yordayıcı olmaktan çıkmıştır. ROC analizi, AH için güçlü bir ayırt edicilik göstermiştir (AUC = 0.89, 95% GA [0.83–0.94]); ≤ 39 kesme noktası için duyarlılık %91 (95% GA [0.71–0.99]) ve özgüllük %82 (95% GA [0.74–0.89]) olarak saptanmıştır. Yaşın modele eklenmesiyle yaşa göre düzeltilmiş modelde tanısal doğruluk daha da artmıştır.

Sonuç: ACT Testi, AH'yi ayırt etmede güçlü bir ayırt edicilik sunmakta ve birden fazla bilişsel alanla anlamlı yakınsak ilişkiler göstermektedir.

Anahtar kelimeler: Alzheimer hastalığı, hafif bilişsel bozukluk, çalışma belleği, nöropsikolojik testler, ROC eğrisi

Address for Correspondence: Ata Cantürk Doğrul, Hacettepe University Department of Psychology, Ankara, Türkiye

Email: atadogrul@hacettepe.edu.tr

Received: 08.10.2025 Accepted: 14.01.2026

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia worldwide and, because of its progressive nature and the lack of a definitive treatment, is a major public health concern¹. Consequently, early diagnosis is crucial in AD, as interventions are most effective during the prodromal stages². In this framework, Mild Cognitive Impairment (MCI) is defined as a transitional stage between normal aging and dementia. Importantly, MCI is a strong predictor of progression to AD. Longitudinal studies indicate an annual conversion rate of about 10–15%^{3,4}. However, distinguishing individuals with MCI or early-stage AD from cognitively healthy older adults remains challenging. Several reviews suggest that conventional screening tools may not detect subtle deficits in attention, working memory (WM), and executive functions⁵. In light of these challenges, the Auditory Consonant Trigram (ACT) Test, adapted from the Brown-Peterson paradigm, assesses WM by preventing rehearsal through a concurrent interference task^{6,7}. In this procedure, participants recall three consonants after delay intervals of 0, 9, 18, or 36 seconds while performing a distractor activity, such as backward counting. Accordingly, the psychometric properties of the ACT Test have been systematically examined across various populations. Early studies reported acceptable test-retest reliability and initially indicated that age did not influence performance⁸.

However, subsequent studies have demonstrated age-related declines in ACT Test performance⁹. In addition, a lower educational background is consistently associated with reduced performance¹⁰. By contrast, most studies have found no significant sex differences among healthy adults^{11,12,13}. More recent evidence has further confirmed the ACT Test's reliability ($\alpha = .79$) and validity in a large sample. These studies reported significant associations with education, processing speed, and WM, as well as sensitivity to various neurological conditions¹¹. Similarly, the Turkish version demonstrated strong internal consistency ($\alpha = .85$) and provided normative data¹². Nevertheless, a university-based study later supported the test's stability ($\alpha = .75$) and found that male participants performed better than female participants at longer delay intervals¹⁴.

Beyond normative research, the ACT Test has shown clinical utility in distinguishing between patient groups and healthy controls. These groups include patients with traumatic brain injury¹⁵, AD^{16,17}, schizophrenia¹², multiple sclerosis¹⁸, and attention deficit hyperactivity disorder¹⁹. In Turkish samples, ACT Test performance showed strong correlations with the Paced Auditory Serial Addition Test (PASAT) among individuals with multiple sclerosis¹⁸. These findings support the ACT Test as a brief and well-tolerated measure of cognitive impairment.

As outlined above, the ACT Test draws on verbal WM demands inherent to the Brown-Peterson paradigm. At the cognitive level, this pattern can be explained within the multicomponent WM model. In this model, the central executive maintains verbal information, suppresses automatic articulatory rehearsal, and manages distraction, thereby increasing reliance on controlled processing^{20,21}. This theoretical framework is reflected in the structure of the ACT Test. In the ACT Test, participants are required to retain consonant trigrams while completing an interference task. This places increased demands on interference resolution and makes performance sensitive to limitations in controlled processing. From a complementary perspective, ACT performance can also be explained by interference-based accounts of short-term retention. In this task structure, retention under distraction becomes particularly vulnerable to proactive and retroactive interference²². Consistent with these characteristics, the ACT Test differs from commonly used WM measures by its explicit sensitivity to interference. Compared with simpler digit span tasks and more computationally demanding measures such as the PASAT, the ACT Test integrates delayed recall, attentional control, dual-task coordination, and resistance to interference within a single paradigm²³.

Empirical evidence indicates that susceptibility to interference increases along the cognitive decline continuum. Interference-related disruptions in WM are evident even at the MCI stage²⁴. Moreover, heightened implicit interference has been detected in cognitively normal older adults who are at increased risk²⁵. Against this background, the ACT Test is particularly sensitive to subtle and progressive interference-related impairments in verbal WM observed in early-stage AD^{26,27,28}.

Given these cognitive and interference-control mechanisms, ACT Test performance is expected to

decline progressively across groups (Control, MCI, AD) and to correlate with other neuropsychological measures. If the ACT Test captures vulnerability to interference, it should also aid diagnostic classification and show clinically meaningful diagnostic accuracy.

Despite its psychometric strengths and established clinical relevance, the diagnostic accuracy of the ACT Test has not been systematically examined in Turkish patients with MCI or AD. This study addresses this gap by evaluating the ACT Test's clinical utility through several methods. Known-groups validity was assessed by comparing diagnostic categories, convergent validity was examined through correlations with other neuropsychological tests, predictive validity was tested using regression analyses, and diagnostic accuracy was evaluated with receiver operating characteristic (ROC) analyses to identify optimal cut-off scores for differentiating clinical groups. This multi-method validation aims to provide comprehensive evidence for the diagnostic usefulness of the ACT Test in Turkish clinical practice.

METHODS

Sample

A priori sample size estimation was conducted using G*Power 3.1²⁹. Because G*Power does not offer a specific option for ROC analyses, a t-test approach based on the expected mean difference between two independent groups was used. The minimum required sample size was 21 participants for the patient group and 105 for the control group ($d = .80$, $\alpha = .05$, one-tailed, $1 - \beta = .95$, allocation ratio $N_2/N_1 = 5$).

Participants in the clinical group were recruited via convenience sampling from individuals who agreed to participate following invitation by a neurologist during routine clinical visits. The control group was recruited using snowball sampling in the same geographical region. In total, 166 volunteers aged 50–86 years were included. The clinical group comprised 42 patients recruited from a private neurology clinic in Ankara. Of these, 22 patients (10 females, 12 males; $M = 68.18$ years) received an early-stage AD diagnosis based on the 2011 National Association on Aging and Alzheimer's Association (NIA-AA)³⁰ criteria following neurological examination. The remaining 20 participants (10 females, 10 males; $M =$

75.05 years) were evaluated for MCI according to the Petersen criteria.³¹ All clinical participants were assessed while receiving their usual medications. The control group consisted of 124 cognitively healthy older adults (70 females, 54 males; $M = 56.29$ years) living independently in the community. Controls were matched to the clinical group by sex and handedness. However, due to the progressive nature of the disease and the older age range of clinical referrals, the clinical groups were not age equivalent (MCI > AD > Control). All participants were native Turkish speakers and had at least a primary school education.

The inclusion criteria were defined separately for each group. The AD group included individuals with Mini-Mental State Examination (MMSE) scores below 24³² and Enhanced Cued Recall Test (ECR Test) total scores below 41³³. The MCI group included participants who reported subjective memory complaints, had a normal neurological examination, and showed objective impairment in at least one cognitive domain (≥ 1.5 SD below age- and education-adjusted norms)³¹. Participants were also required to have preserved activities of daily living and MMSE scores of 24 or higher. The healthy control group included individuals with MMSE scores of 24 or higher and normal performance on standardized cognitive tests. Exclusion criteria for all groups included scores above 14 on the Geriatric Depression Scale (GDS)³⁴, focal neurological signs, focal lesions on neuroimaging (CT or MRI), a history of significant head trauma, and any psychiatric or neurological disorder that could affect cognitive performance. Individuals with uncorrected visual or auditory impairments were also excluded.

Participant recruitment for the clinical and control groups, as well as participant flow across analytical stages, are summarized in Figure 1, which was prepared in accordance with the STARD 2015 guidelines for diagnostic accuracy studies³⁵. Demographic characteristics of the sample are presented in Table 1.

Measures

All neuropsychological tests and scales used in the study are widely used in routine clinical practice in Türkiye and have shown good reliability and validity in previous standardization and validation studies. Accordingly, all measures exhibit strong psychometric properties.

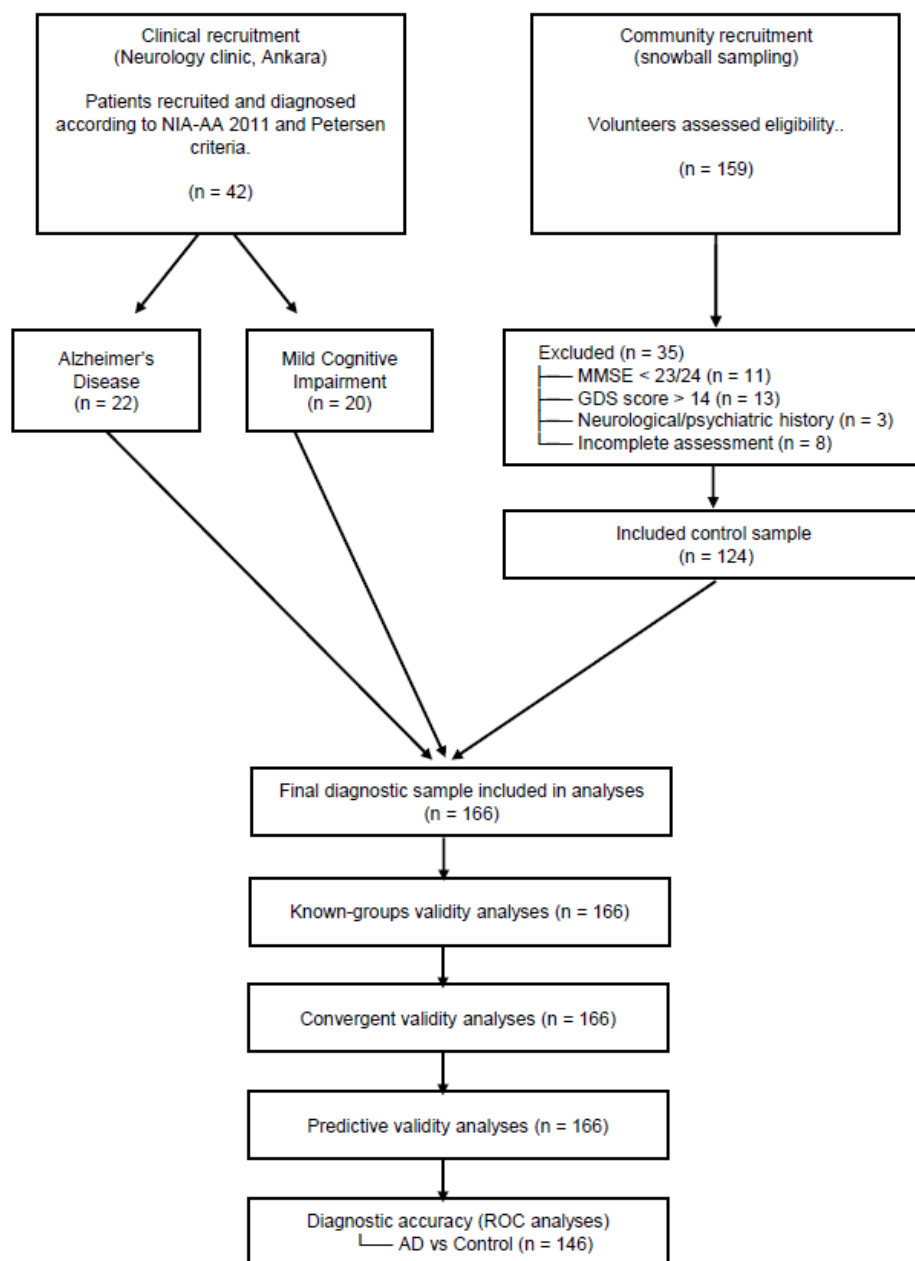


Figure 1. Participant flow diagram

Prepared in accordance with STARD 2015³⁵ recommendations for reporting diagnostic accuracy studies.

The clinical sample met the predefined inclusion and exclusion criteria (including $GDS \leq 14$ and absence of major psychiatric or neurological comorbidity); however, the total number of patients screened for eligibility was not systematically recorded.

MMSE = Mini-Mental State Examination; GDS = Geriatric Depression Scale; ROC = receiver operating characteristic.

Table 1. Comparison of demographic variables across groups.

Variable	AD (n = 22)	MCI (n = 20)	Control (n = 124)	Statistics	p
Age (<i>M</i> ± <i>SD</i>)	68.18 ± 5.09	75.05 ± 6.04	56.29 ± 5.95	$H_{(2)} = 74.74$	< .001
Sex (Female/Male)	12 (54.5%)/ 10 (45.5%)	10 (50.0%)/ 10 (50.0%)	54 (43.5%) / 70 (56.5%)	$\chi^2_{(2)} = 1.07$	> .05
Education level (n, %)				$\chi^2_{(8)} = 12.97$	> .05
Primary school	2 (9.0%)	2 (10.0%)	20 (16.1%)		
Secondary school	1 (4.6%)	0 (0.0%)	11 (8.9%)		
High school	6 (27.3%)	3 (15.0%)	22 (17.7%)		
Associate degree	1 (4.6%)	5 (25.0%)	10 (8.1%)		
Bachelor's degree +	12 (54.6%)	10 (50.0%)	61 (49.2%)		

Age was analyzed with the Kruskal–Wallis (H) test; sex and education level were analyzed with Chi-square (χ^2) tests.; *p* value < .05 was considered statistically significant.; *M* = Mean; *SD* = Standard Deviation; AD = Alzheimer's disease; MCI = Mild cognitive impairment.

Mini-Mental State Examination

The Mini-Mental State Examination (MMSE) is a widely used brief cognitive screening instrument consisting of 30 items that assess orientation, attention, immediate and delayed recall, language, and visuospatial abilities. The total score ranges from 0 to 30, with lower scores indicating greater cognitive impairment³⁶.

The MMSE has been adapted and validated for the Turkish population and has demonstrated strong psychometric properties. A cut-off score of 23/24 has been recommended, with scores of 23 or below indicating a high likelihood of cognitive impairment. At this threshold, the MMSE showed sensitivity of .91, specificity of .95, and a kappa value of .86 for distinguishing patients with dementia from healthy controls. Inter-rater reliability was excellent (reliability coefficient of .99; kappa value of .92)³⁴. Administration of the MMSE takes about 10 minutes.

Enhanced Cued Recall Test

The Enhanced Cued Recall (ECR) Test is a widely used measure that distinguishes the memory performance of healthy older adults from patients with dementia³⁷. The test has been adapted and validated for the Turkish population, supporting its use in both clinical and research settings³³.

In the Turkish validation study, a total recall cut-off score of 41/48 distinguished patients with AD from healthy controls with 100% sensitivity and 93.9% specificity (AUC = 0.99, 95% CI [0.93–0.99]). The same cut-off distinguished dementia from non-demented individuals with 91.8% sensitivity and 80.8% specificity (AUC = 0.91, 95% CI [0.83–0.95]). Although the test is less effective for distinguishing MCI from healthy controls, the total recall score demonstrated moderate diagnostic value (cut-off =

42, sensitivity = 50.0%, specificity = 90.9%, AUC = 0.62).

The test materials consist of 16 black-and-white line drawings presented in a booklet comprising four cards, each containing four items. During the encoding phase, participants view each card and name the drawings using a provided semantic cue to ensure controlled encoding. Following encoding, three recall trials are administered, each including free recall followed by cued recall for items not retrieved spontaneously. The procedure takes about 15 minutes.

Trail Making Test

The Trail Making Test (TMT) is a widely used neuropsychological measure of attention, visual scanning, processing speed, and executive functioning³⁸. The test consists of two parts: TMT-A and TMT-B.

In TMT-A, participants connect numbered circles in ascending order (1-2-3, etc.), primarily assessing visual search and psychomotor speed. In TMT-B, participants alternate between numbers and letters (1-A-2-B-3-C, etc.), which requires greater cognitive flexibility and set-shifting. Performance is scored by completion time, with longer times indicating greater cognitive difficulty. Each part provides three indices: completion time, errors, and corrections. To reduce the overall duration of the neuropsychological assessment, only TMT-B was administered. For the present analyses, only TMT-B completion time was included, as error and correction indices are highly correlated with completion time and may introduce multicollinearity in regression models.

The Turkish adaptation closely follows the original format, with minor changes to the alphabet. A large-scale normative study of adults aged 50 and older

established age- and education-stratified norms and showed strong reliability, with test-retest coefficients of $r = .78$ for TMT-A and $r = .73$ for TMT-B. Interrater reliability was also high ($r = .99$ for TMT-A; $r = .93$ for TMT-B)³⁹. In a later Turkish clinical validation study, TMT-B showed good diagnostic accuracy in distinguishing patients with AD from healthy older adults, with a cut-off score of 240 seconds yielding a sensitivity of .84 and a specificity of .90⁴⁰. Administration of the TMT takes about 10 minutes.

Clock Drawing Test

The Clock Drawing Test (CDT), first introduced as part of the Boston Aphasia Battery, is a widely used neuropsychological screening tool that assesses comprehension, planning, visuospatial abilities, and executive functioning⁴¹. Clinically, it offers a brief and practical measure of global cognition and is especially useful for distinguishing healthy aging from dementia⁴².

Several scoring systems exist; in this study, the four-point scoring method was used. Participants were asked to draw a clock face without a pre-drawn circle and set the time to "11:10." The Turkish validation study showed good reliability, with a test-retest correlation coefficient of $r = .88$ and an inter-rater reliability coefficient of $r = .74$, indicating adequate temporal stability and scoring consistency⁴². Administration of the CDT takes about 5 minutes.

Stroop Test (TBAG Version)

The Stroop Test is a widely used measure of executive functions, assessing cognitive interference, selective attention, and processing speed. The TBAG version of the Stroop Test, adapted into Turkish from the Victoria version, includes five subtests scored by completion time, errors, and corrections, with a maximum total of 15 points. The Stroop Test (TBAG version) provides indices of executive functioning, attention, and cognitive flexibility, and has been validated for use in the Turkish population⁴³.

The Turkish adaptation study reported strong reliability, with a test-retest correlation of $r = .83$ for the interference score and inter-rater reliability coefficients above $r = .94$, indicating high scoring consistency.

In the present study, only the completion time for the Color-word interference subtest (Stroop-5) was used, as this measure is considered the most sensitive indicator of executive functioning. Error and

correction indices are highly correlated with completion time, so including them would increase the risk of multicollinearity; therefore, only the time variable was analyzed. The procedure takes about 10 minutes.

Auditory Consonant Trigram Test

The Auditory Consonant Trigram (ACT) Test is a paper-and-pencil measure that assesses WM, divided attention, and information processing^{6,7,12}. The test includes 20 items on a single A4 sheet. For the first five items, participants recall three consonants immediately after presentation, with a zero-second delay. For the remaining 15 items, each consonant trigram is followed by a two- or three-digit number. Participants then count backward aloud for 3, 9, or 18 seconds before recalling the letters in any order.

Subtest scores indicate the number of correctly recalled letters at each delay interval (0, 3, 9, 18 seconds), and the total score is the sum of all subtests (range: 0–60). Turkish normative data for the ACT Test have been established in two studies. The first study collected data from 236 healthy individuals aged 16–65 with varying educational backgrounds and reported high internal consistency ($\alpha = .85$), as well as validity based on group differences between schizophrenia patients and matched controls, and correlations with the Backward Digit Span Test¹². A more recent study updated the norms in a younger cohort of 304 university students aged 18–26, providing sex-specific reference values. This study also reported acceptable internal consistency for the total score ($\alpha = .85$) and found performance differences across delay intervals, with males performing better at longer delays¹⁴. Administration requires about 10 minutes.

Geriatric Depression Scale

The Geriatric Depression Scale (GDS) is a self-report screening tool designed to assess depressive symptoms in older adults⁴⁴. The original version includes 30 yes/no items that address mood, cognitive complaints, and functional aspects of depression; shorter forms, such as 15-item versions, are also available. In this study, researchers administered the 30-item long form. The dichotomous response format improves feasibility, especially for older individuals and those with MCI and AD.

The Turkish adaptation and validation of the GDS showed strong psychometric properties, with internal

consistency of $\alpha = .92$ and test-retest reliability of $r = .77^{34}$. The GDS is widely used in clinical and research settings as a practical and reliable measure of late-life depression. Administration takes about 5 minutes.

Procedure

Ethical approval was obtained from the Hacettepe University Health Sciences Research Ethics Committee (research number: SBA 25/220; date: 04/03/2025; decision number: 2025/08-43). Written informed consent was obtained from all healthy control participants and from caregivers or legal representatives of participants in the patient groups before enrollment. Clinical diagnoses of MCI and AD were established by a neurologist according to internationally accepted criteria (NIA-AA 2011³⁰, Petersen³¹).

The study procedures were conducted in the laboratories of the Department of Psychology at Hacettepe University for the healthy control group, whereas assessments for the patient groups were carried out during routine clinical visits at a private neurology clinic in Ankara. Neuropsychological assessments were administered by a trained psychologist across two sessions. The psychologist was a member of the research team and had formal training in neuropsychological assessment.

In the first session, participants completed the GDS, MMSE, ECR Test, TMT-B, CDT, and Stroop-5. The order of tests was randomized to control for sequence effects. In the second session, participants completed the ACT Test. Participant alertness was monitored throughout both sessions, and short rest breaks were provided when needed. A scheduled 10-minute break separated the two sessions to reduce fatigue and support engagement during the ACT Test, which was the study's primary measure and required the greatest WM and interference control. The entire assessment protocol lasted approximately 75 minutes.

All data were recorded using anonymized participant codes, and data entry and statistical analyses were performed by the research team.

Statistical analysis

Statistical analyses followed five sequential steps aligned with the study's hypotheses and analytical objectives. Normality of continuous variables was assessed using the Shapiro–Wilk test, and

homogeneity of variances was examined with Levene's test. Variables that violated normality assumptions were analyzed using nonparametric methods.

First, demographic characteristics were examined to determine whether group differences, particularly in age, could influence subsequent cognitive comparisons. Age differences were tested using the Kruskal–Wallis test. Sex, education level, and hand preference were evaluated using chi-square analyses. Second, known-groups validity was examined by comparing neuropsychological test performance, including the ACT Test, across diagnostic groups. This analysis tested the hypothesis of a graded cognitive decline (Control > MCI > AD). Because normality assumptions were not met, Kruskal–Wallis tests were used, followed by Dunn–Bonferroni–corrected pairwise comparisons.

Third, convergent validity was assessed by examining associations between ACT Test performance and other cognitive measures. As several variables deviated from normality and age differed significantly between groups, age-adjusted partial Spearman correlations were calculated between the ACT Test total score and other neuropsychological test scores. Fourth, predictive validity was evaluated using multinomial logistic regression analyses. These models tested whether ACT Test performance predicted diagnostic group membership, both independently and in combination with other neuropsychological measures.

Finally, the diagnostic accuracy of the ACT Test was examined using ROC analyses. Optimal cut-off scores were determined using the Youden Index⁴⁵. Diagnostic indices, including area under the curve (AUC), sensitivity, specificity, likelihood ratios, positive predictive value, and negative predictive value, were reported for both raw and age-adjusted models. Age-adjusted ROC curves were based on predicted probabilities derived from logistic regression models including age and ACT Test total score as joint predictors. All analyses were conducted using IBM SPSS Statistics® (Version 29) and JASP® (Version 0.18).

RESULTS

Table 1 summarizes the demographic characteristics of the sample. Because the age variable did not meet the normality assumption, age differences were examined using a nonparametric Kruskal–Wallis test.

The analysis revealed a significant group effect for age, $H_{(2)} = 74.74, p < .001$. The control group was significantly younger than both the AD and MCI groups. No significant differences were observed for sex distribution or education level, as indicated by Chi-square tests ($p > .05$). Thus, the groups were similar in education and sex but differed significantly in age. Hand preference was consistent across groups, as all participants were right-handed.

Known-groups validity: discrimination of groups based on neuropsychological test measures

Because the normality assumptions were violated for all neuropsychological variables, group differences were examined using Kruskal–Wallis tests. When omnibus effects were significant, post-hoc comparisons were conducted using the Dunn–Bonferroni correction. Table 2 presents the Kruskal–Wallis results for ACT Test scores, and Table 3 summarizes the results for the remaining neuropsychological measures.

ACT test scores

Significant group differences were observed for all delayed recall conditions of the ACT Test, whereas immediate recall did not differ across groups. For the ACT Test total score, the Kruskal–Wallis test revealed a strong group effect, $H_{(2)} = 36.72, p < .001, \eta^2 = 0.21$.

Patients with AD performed significantly worse than both patients with MCI and controls, and patients with MCI scored lower than controls. A similar pattern was found for the 3-second delay condition, with a strong group effect ($H_{(2)} = 35.61, p < .001, \eta^2 = 0.20$). Post-hoc analyses indicated significant differences among all three groups.

The 9-second delay condition a large group effect was observed ($H_{(2)} = 46.91, p < .001, \eta^2 = 0.27$). Both AD and MCI groups recalled fewer items than controls, but the two clinical groups did not differ significantly. A similarly large effect appeared in the 18-second delay condition ($H_{(2)} = 50.84, p < .001, \eta^2 = 0.30$), with patients with AD and MCI again performing worse than controls, and no significant difference between the two clinical groups. In contrast, performance in the 0-second (immediate recall) condition did not differ across groups ($H_{(2)} = 0.45, p > .05, \eta^2 = 0.00$), indicating that preserved immediate recall irrespective of diagnostic status.

Overall, ACT Test performance showed a clear gradient of impairment (Control > MCI > AD) that became more pronounced as the delay interval increased. Differences between the MCI and AD groups diminished at longer delays, with no significant separation between these groups in the later condition. Group means across delay intervals are presented in Figure 2.

Table 2. Kruskal-Wallis test results for ACT Test scores across groups

N = 166	AD (n = 22) M (SD)	MCI (n = 20) M (SD)	Control (n = 124) M (SD)	H(2)	p	η^2	Post-hoc Comparisons
ACT Test total score	33.09 (5.61)	41.22 (5.26)	45.13 (9.57)	36.72	< .001	.21	AD < MCI < Control
ACT Test 0-sec	14.50 (1.14)	14.44 (.73)	14.14 (1.65)	0.45	.80	.00	n.s.
ACT Test 3-sec	8.95 (2.79)	12.11 (2.15)	12.61 (2.35)	35.61	< .001	.20	AD < MCI < Control
ACT Test 9-sec	6.27 (2.68)	8.44 (2.40)	11.81 (2.40)	46.91	< .001	.27	AD = MCI < Control
ACT Test 18-sec	4.64 (1.40)	5.11 (3.26)	8.53 (3.56)	50.84	< .001	.30	AD = MCI < Control

p value < .05 was considered statistically significant.; Effect sizes are reported as eta squared (η^2), with larger values indicating stronger group effects; Higher ACT Test scores indicate better performance. ; Post-hoc comparisons were performed using Dunn–Bonferroni correction.; M = Mean; SD = Standard Deviation; AD = Alzheimer's disease; MCI = Mild Cognitive Impairment; ACT Test = Auditory Consonant Trigram Test. n.s. = non-significant.

Other neuropsychological test scores

Significant group differences were observed across all additional neuropsychological measures (Table 3). For the MMSE, the Kruskal–Wallis test revealed a strong group effect, $H_{(2)} = 45.42, p < .001, \eta^2 = 0.27$. Patients with AD scored significantly lower than both patients with MCI and controls, and patients with

MCI also performed significantly below controls. A similar pattern was found for the ECR Test total score, $H_{(2)} = 64.07, p < .001, \eta^2 = 0.38$. Controls scored significantly higher than both clinical groups, and patients with MCI scored significantly higher than patients with AD, with all pairwise comparisons significant.

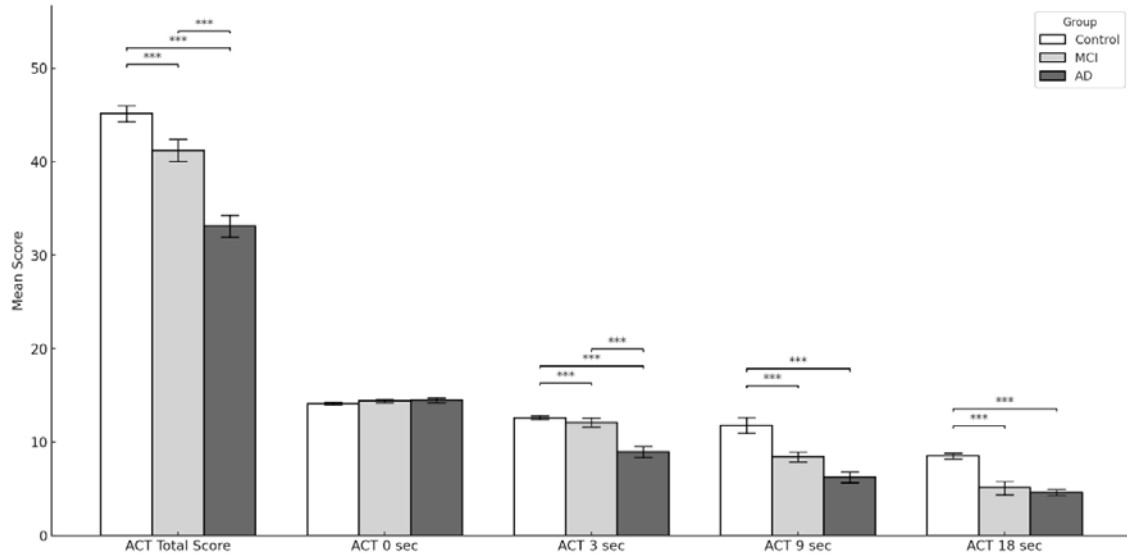


Figure 2. Mean ACT Test performance across groups with standard error bars.

The figure presents mean scores for the ACT Test Total Score and the 0-, 3-, 9-, and 18-second delay conditions for the Control, MCI, and AD groups; *** $p < .001$; Error bars indicate standard errors of the mean; Pairwise differences based on Dunn–Bonferroni post-hoc tests.

Table 3. Kruskal-Wallis test results for other neuropsychological test scores.

N = 166	AD (n = 22) M (SD)	MCI (n = 20) M (SD)	Control (n = 124) M (SD)	$H_{(2)}$	p	η^2	Post-hoc Comparisons
MMSE	21.18 (3.25)	26.50 (2.21)	27.27 (1.82)	45.42	< .001	.27	AD < MCI < Control
ECR Test	29.45 (9.55)	41.45 (5.62)	46.23 (1.94)	64.07	< .001	.38	AD < MCI < Control
CDT	3.27 (1.35)	3.90 (0.45)	3.69 (.63)	8.53	< .05	.04	AD < MCI, Control
Stroop-5	59.09 (45.05)	50.89 (23.14)	31.18 (25.51)	40.76	< .001	.24	AD = MCI > Control

p value < .05 was considered statistically significant; Effect sizes are reported as eta squared (η^2), with larger values indicating stronger group effects; Stroop-5 and TMT-B reflect completion time scores; higher values indicate poorer performance; Post-hoc comparisons were performed using Dunn–Bonferroni correction; M = Mean; SD = Standard Deviation; AD = Alzheimer’s disease; MCI = Mild Cognitive Impairment; MMSE = Mini-Mental State Examination; ECR Test = Enhanced Cued Recall Test; CDT = Clock Drawing Test; Stroop-5 = Color–word interference subtest; TMT-B = Trail Making Test–Part B

Group differences were observed for the CDT, although with a smaller effect size, $H_{(2)} = 8.53, p < .05, \eta^2 = 0.04$. Patients with AD scored lower than both patients with MCI and controls, whereas the MCI and control groups did not differ. For the Stroop-5 completion time, a significant group effect was found, $H_{(2)} = 40.76, p < .001, \eta^2 = 0.24$. Both patients with AD and those with MCI took longer than controls, but there was no significant difference between the AD and MCI groups.

Finally, the TMT-B completion time showed a robust group effect, $H_{(2)} = 27.60, p < .001, \eta^2 = 0.16$. Patients with AD took significantly longer to complete the task than both patients with MCI and controls, and patients with MCI also required significantly more time than controls. Overall, performance on global cognition, memory, and executive functioning followed a consistent gradient of impairment (Control > MCI > AD), paralleling the pattern observed for ACT Test performance.

Convergent validity: associations between ACT test and other neuropsychological test scores

Because all variables violated normality assumptions, non-parametric procedures were used. To account for significant age differences between groups, age-adjusted partial Spearman correlations were calculated. After controlling for age, ACT Test total scores had moderate positive associations with general cognition (MMSE; $r = .49$, $p < .001$) and memory performance (ECR Test; $r = .42$, $p < .001$). ACT Test scores were negatively associated with processing speed and executive functioning, as measured by Stroop-5 ($r = -.40$, $p < .001$) and TMT-B completion times ($r = -.29$, $p < .001$). No association was found between ACT Test performance and CDT scores ($r = .03$, $p > .05$).

To examine whether these associations differed across diagnostic groups independent of between-group variance, additional Spearman's rho correlations were computed separately for the AD, MCI, and control groups (Table 4). In the AD group,

ACT Test scores were strongly related to MMSE ($r = .61$, $p < .001$) and ECR Test performance ($r = .53$, $p < .05$). In the MCI group, these relationships were moderate ($r = .38-.44$), while in the control group, associations were modest ($r = .18$ and $.34$).

Similarly, negative associations between Stroop-5 and TMT-B completion times were strongest in AD ($r = -.49$ and $-.41$), smaller in mild cognitive impairment ($r = -.35$ and $-.33$), and weakest or nonsignificant in controls ($r = -.23$ and $-.16$). Across all groups, ACT Test performance showed no meaningful association with CDT scores, indicating that visuoconstruction skills were largely independent of ACT Test performance.

The associations between ACT Test performance and all cognitive measures became stronger from healthy controls to the AD group. MMSE and ECR Test showed progressively higher positive correlations, while Stroop-5 and TMT-B showed increasingly stronger negative correlations. CDT remained weakly related to ACT Test in all groups (Table 4).

Table 4. Spearman correlations between ACT Test total score and other neuropsychological test scores.

N = 166 Neuropsychological Tests	AD (n = 22)		MCI (n = 20)		Control (n = 124)	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
MMSE total score	.61	< .001	.44	< .05	.34	< .001
ECR Test total score	.53	< .05	.38	< .05	.18	< .05
CDT total score	.05	> .05	.06	> .05	-.01	> .05
Stroop-5	-.49	< .05	-.35	< .05	-.23	< .05
TMT-B	-.41	< .05	-.33	< .05	-.16	> .05

p value < .05 was considered statistically significant; Negative correlations for Stroop-5 and TMT-B reflect time-based scoring (higher times = poorer performance); Correlations were computed separately within groups using Spearman's *r* due to non-normal distributions. All tests were two-tailed; AD = Alzheimer's disease; MCI = Mild Cognitive Impairment; MMSE = Mini-Mental State Examination; ECR Test = Enhanced Cued Recall Test; CDT = Clock Drawing Test; Stroop-5 = Color-word interference subtest; TMT-B = Trail Making Test-Part B.

Predictive validity: multinomial logistic regression analysis

Multinomial logistic regression analyses were used to assess the predictive validity of the ACT Test total score (Table 5). The linearity of the logit was tested using the Box-Tidwell procedure; after Bonferroni correction, no significant violations were found. Influence diagnostics identified one case with Cook's distance greater than 1 and a studentized residual of |3.0|, but removing this case did not meaningfully change parameter estimates. Model fit was assessed with the likelihood ratio chi-square test, Nagelkerke R^2 , and classification accuracy.

In the first model, which included age and ACT Test total score, model fit was good, $\chi^2(4) = 86.34$, $p < .001$, with Nagelkerke $R^2 = .45$ and 85.8% classification accuracy. Age significantly predicted both AD (OR = 1.25, 95% CI [1.14–1.37], $p < .001$) and MCI (OR = 1.41, 95% CI [1.20–1.65], $p < .001$). The ACT Test total score significantly distinguished AD from controls (OR = 0.89, 95% CI [0.84–0.96], $p < .05$) but did not distinguish MCI ($p > .05$).

In the second model, which included MMSE, ECR Test, CDT, Stroop-5, and TMT-B, model fit improved substantially, $\chi^2(14) = 121.73$, $p < .001$, with Nagelkerke $R^2 = .82$ and 90.3% classification accuracy.

In this expanded model, the ACT Test total score was not a significant predictor ($p > .05$). The ECR Test was the only consistent cognitive predictor for both AD (OR = 0.39, 95% CI [0.16–0.96], $p < .05$) and MCI (OR = 0.34, 95% CI [0.13–0.96], $p < .05$). Age remained a significant predictor for both AD ($p < .05$) and MCI ($p < .01$), whereas no other measures improved classification.

Overall, the analyses indicated that age consistently predicted both AD and MCI status, the ACT Test total score differentiated AD from controls but not MCI, and its independent predictive value was attenuated when other neuropsychological measures were included in the model.

In Model 1 (age and ACT Test), multicollinearity was evident, with high VIF values (27–47) and low tolerance ($< .04$) values, reflecting a strong correlation between age and ACT Test in this clinical sample. In Model 2, similarly high VIF values (27–136) and low tolerance values ($< .04$) were observed, indicating substantial shared variance among ACT Test, MMSE, and ECR Test, which assess overlapping age-related cognitive domains. Although the ACT Test contributes to the overall diagnostic performance of the models, its independent predictive value decreases when multicollinearity among neuropsychological measures is considered.

Table 5. Multinomial Logistic Regression Results and Multicollinearity Diagnostics

Predictors	AD vs Control OR [95% CI]	<i>p</i>	MCI vs Control OR [95% CI]	<i>p</i>	VIF	Tolerance
Model 1: Age + ACT Test						
Age	1.25 [1.14-1.37]	< .001	1.41 [1.20-1.65]	< .001	47.85	0.02
ACT Test	0.89 [0.84-0.96]	< .05	1.04 [.91-1.19]	> .05	27.18	0.04
Model Fit	$\chi^2_{(4)} = 86.34, p < .001, R^2 = .45, Accuracy = 85.8\%$					
Model 2: Age + ACT Test + Other Neuropsychological Tests						
Age	1.79 [1.11-2.87]	< .05	2.06 [1.11-3.84]	< .05	47.85	0.02
ACT Test	0.76 [0.45-1.28]	> .05	1.43 [0.87-2.35]	> .05	27.18	0.04
MMSE	0.19 [0.03-1.19]	> .05	0.46 [0.17-1.19]	> .05	136.11	0.01
ECR Test	0.39 [0.16-0.96]	< .05	0.34 [0.13-0.96]	< .05	84.62	0.01
CDT	36.8 [0.03-3.9 × 10 ⁴]	> .05	8.17 [0.35-191.3]	> .05	25.88	0.04
Stroop-5	0.90 [0.74-1.09]	> .05	1.03 [0.99-1.07]	> .05	2.79	0.36
TMT-B	1.11 [0.95-1.30]	> .05	1.01 [0.97-1.05]	> .05	6.52	0.15
Model Fit	$\chi^2_{(14)} = 121.73, p < .001, R^2 = .82, Accuracy = 90.3\%$					

p value < .05 was considered statistically significant; OR with 95% confidence intervals CI are reported. OR > 1 indicates increased odds of belonging to the clinical group relative to controls, whereas OR < 1 indicates decreased odds; Higher scores on ACT Test, MMSE, ECR Test, and CDT indicate better performance; Stroop-5 and TMT-B are completion time measures, with higher scores indicating poorer performance; Model fit is reported using likelihood ratio χ^2 , Nagelkerke R^2 , and classification accuracy.; Multicollinearity diagnostics are presented as VIF and tolerance; AD = Alzheimer’s disease; OR = Odds Ratio; CI = Confidence Interval; VIF = Variance Inflation Factor; ACT Test= Auditory Consonant Trigram Test; MMSE = Mini-Mental State Examination; ECR Test = Enhanced Cued Recall Test; CDT = Clock Drawing Test; Stroop-5 = Color–word interference subtest; TMT-B = Trail Making Test–Part B.

Diagnostic accuracy and ROC analyses

To assess the diagnostic utility of the ACT Test total score, ROC analyses distinguished individuals with AD from healthy controls. Consistent with the regression analysis ($p > .05$), ACT Test performance did not significantly differentiate the MCI group from the control group. Therefore, a separate ROC analysis for the MCI group was considered statistically unreliable and clinically ambiguous, so ROC analyses were conducted only for the AD group.

Because higher ACT Test scores indicate better performance, raw ROC curves were computed using reversed ACT Test scores so that higher values reflected a greater likelihood of impairment. Optimal cut-off values were determined using the Youden Index⁴⁵ for the AD group. Because there were significant age differences between groups, age-adjusted ROC curves were also derived by fitting logistic regression models with both ACT Test total score and age as predictors and using the resulting predicted probabilities for ROC computation. This approach allowed evaluation of the ACT Test’s

diagnostic accuracy independent of age-related effects.

The raw ROC analysis showed excellent discrimination (AUC = 0.89, 95% CI [0.83–0.94]), with a Youden-derived optimal cut-off score of ≤ 39 , sensitivity of 0.91 (95% CI [0.71–0.99]), and specificity of 0.82 (95% CI [0.74–0.89]) (Table 6). The age-adjusted model further improved diagnostic performance (AUC = 0.95, 95% CI [0.91–0.98]); the updated optimal probability cut-off of $p \geq .143$ yielded sensitivity of 1.00 (95% CI [0.85–1.00]) and specificity of 0.89 (95% CI [0.82–0.94]) (see Figure 3). DeLong's test showed that the improvement from the raw to the age-adjusted model was statistically significant ($z = 2.48$, $p < .05$), confirming that adjusting for age substantially improved diagnostic accuracy.

To further enhance the clinical interpretability of ROC findings, likelihood ratios (LR+ and LR-) and positive and negative predictive values (PPV and NPV) were calculated for the raw and age-adjusted models. Subsequently, alternative cut-off values were derived based on predefined sensitivity- and specificity-oriented criteria and are presented in Table 6.

In addition to sensitivity and specificity, likelihood ratios and predictive values were calculated to improve clinical interpretability. The raw model produced LR+ = 5.14 and LR- = .11, with a positive predictive value (PPV) of .48 and a negative predictive value (NPV) of .98 at the observed prevalence of 15.1%. The age-adjusted model showed LR+ = 8.86, LR- = .00, PPV = .61, and NPV = 1.00, indicating very high discriminative power for distinguishing AD from healthy controls

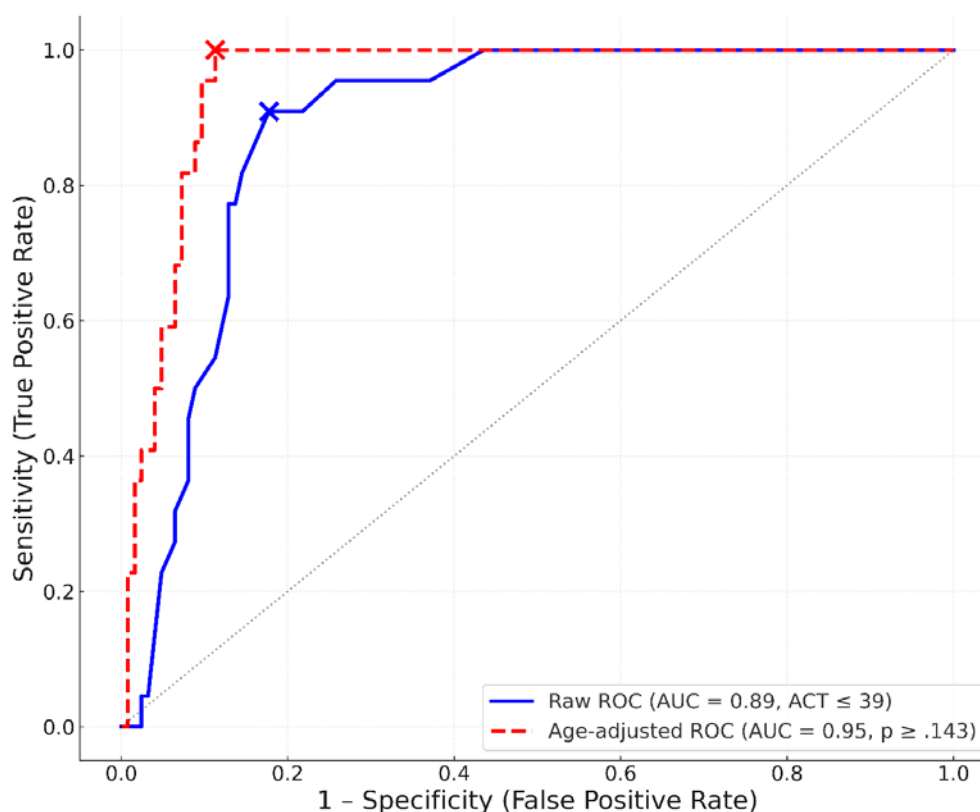


Figure 3. ROC curves for the ACT Test total score in distinguishing AD from healthy controls.

The "x" markers indicate the points of maximal discrimination in the sensitivity–specificity space, as determined by the Youden Index⁴⁵; the corresponding values are provided in the legend; AUC = Area Under the Curve; ROC = Receiver Operating Characteristic.

Table 6. Diagnostic accuracy indices for the ACT Test total score in distinguishing AD from healthy controls.

Group Comparison	AUC	95% [CI]	SE	Cut-off	Se	Sp	LR +	LR -	PPV	NP V	Covariate Adjustment
AD vs Control	0.89	[0.83-0.94]	.03	≤ 39	.91	.82	5.14	.11	.48	.98	None
AD vs Control	0.95	[0.91-0.98]	.02	$p \geq .143$	1.00	.89	8.86	.00	.61	1.00	Age-adjusted*
Alternative Cut-off Values											
Group Comparison	Se / Sp profiles		Cut-off	Se	Sp	LR +	LR -	PPV *	NP V	Covariate Adjustment	
AD vs Control	High Se / Lower Sp		≤ 41	.95	.74	3.70	.06	.40	.99	None	
AD vs Control	High Sp / Lower Se		≤ 33	.50	.91	5.64	.55	.50	.91	None	
AD vs Control	Very high Sp / Low Se		≤ 27	.23	.95	4.70	.81	.45	.87	None	

Higher ACT Test total scores indicate better performance; lower scores indicate greater impairment; PPV and NPV were calculated using the observed AD prevalence in the sample (15.1%); High sensitivity and high/very high specificity thresholds were defined a priori using conventional criteria (Se ≥ .95; Sp ≥ .90 and Sp ≥ .95, respectively); For the age-adjusted model, the cut-off value ($p \geq .143$) represents the optimal predicted probability threshold derived from the logistic regression model including both ACT Test total score and age, rather than a raw ACT Test score. DeLong’s test comparing raw and age-adjusted ROC curves indicated a significant difference in AUC values ($\chi = 2.48, p < .05$), confirming that adjusting for age significantly improved diagnostic accuracy; AD = Alzheimer’s disease; AUC = Area Under the Curve; CI = Confidence Interval; SE = Standard Error; Se = Sensitivity; Sp = Specificity; LR+ = Positive likelihood ratio; LR- = Negative likelihood ratio; PPV = Positive predictive value; NPV = Negative predictive value.

DISCUSSION

This study examined the diagnostic value of the ACT Test using a multi-method, comprehensive validity approach in a Turkish clinical sample including patients with AD, MCI, and healthy older adults. Overall, the findings indicate that the ACT Test is a valid and clinically informative measure. It appears sensitive to impairments in verbal WM, particularly under conditions of interference, which commonly decline in AD.

Before interpreting cognitive outcomes, demographic comparability was examined. Although the groups differed in age and statistical adjustment was applied, neuropsychological tests inherently measure age-sensitive components, such as processing speed, attentional stability, and WM efficiency⁴⁶. Because a substantial portion of age-related variance is shared across cognitive measures, statistical adjustment cannot fully remove residual age effects⁴⁷. Therefore, some non-specific age influences may have contributed to the observed group differences^{46,47}.

Within this context, analysis of ACT performance revealed a clear delay-dependent gradient of impairment, consistent with the total ACT Test score pattern (AD < MCI < Control). The 0-second condition did not distinguish the groups, likely

because immediate recall requires minimal WM. When interference was introduced, group differences became clear: at 3 seconds, AD participants performed significantly worse than both MCI and controls, while at 9 and 18 seconds, AD and MCI groups showed similar performance, but both remained well below controls.

This convergence at longer delays may indicate increased task difficulty, floor effects, or reduced between-group variability, rather than a mechanism driven solely by interference. Because these alternative explanations were not formally tested, delay-dependent patterns should be interpreted cautiously. Nevertheless, this finding is consistent with prior evidence showing that interference disproportionately disrupts short-term retention in AD^{16,17}. Using a related Brown–Peterson paradigm, Morris similarly reported pronounced forgetting under interference despite methodological differences¹⁷. Importantly, these findings do not directly indicate that the ACT Test isolates a single cognitive mechanism; rather, performance likely reflects the combined influence of multiple age- and disease-sensitive processes. Consistent with this interpretation, a comparable impairment gradient (AD < MCI < Control) was observed across MMSE, ECR Test, Stroop-5, and TMT-B. Supporting this view, Shura et al. found that ACT Test performance

depends on short-term memory, processing speed, attentional control, and WM updating¹¹.

The correlation analyses provided additional insight into the cognitive domains associated with the ACT Test and offered evidence for its convergent validity. ACT Test scores showed moderate associations with general cognitive functioning (MMSE) and long-term and recognition memory (ECR Test) across groups, with stronger associations observed in the AD group. This pattern may indicate that ACT Test performance becomes more closely aligned with overall cognitive status as neurodegeneration progresses, although causal interpretations cannot be drawn. By contrast, associations with the CDT were weak, suggesting limited involvement of visuospatial construction. In the AD and MCI groups, negative associations with Stroop-5 and TMT-B were observed. This pattern suggests a potential contribution of processing speed and executive control, which is consistent with the interference-based demands of the task.

To evaluate the predictive validity of ACT Test performance, multinomial logistic regression analyses were conducted. In the model including age and ACT Test scores, both variables significantly predicted diagnostic status. ACT Test distinguished AD from controls but did not differentiate MCI from controls. This pattern is consistent with evidence that interference-resistant working-memory deficits are more pronounced in AD than in prodromal stages^{26,48}. Age was also a strong predictor of both clinical conditions, reflecting its role as a major risk factor⁴⁷. The lack of ACT Test discrimination for the MCI group should be considered in the context of the clinical profile of the current sample. In this sample, ECR Test performance was relatively high ($M = 41.45$, $SD = 5.62$), indicating preserved episodic memory functioning.

This observation is consistent with a previous Turkish validation study reporting modest diagnostic accuracy of the ECR Test for distinguishing MCI from controls ($AUC = .62$)³³.

However, although the ACT Test primarily assesses interference-resistant verbal WM, performance also partly depends on long-term memory processes mediated by the episodic buffer²¹. The moderate correlation between ACT Test and ECR Test scores suggests that these tasks likely share overlapping cognitive components. This overlap may reduce the

ability to distinguish the MCI group when episodic memory is relatively preserved.

When additional neuropsychological tests (MMSE, ECR Test, CDT, Stroop-5, TMT-B) were included, the unique predictive value of the ACT Test decreased. This attenuation reflects the substantial multicollinearity among cognitive variables. ACT Test performance shares significant variance with episodic memory and executive function measures, which are domains that decline early and markedly in AD^{49,50}. High multicollinearity among cognitive measures is a common and expected feature of multivariate neuropsychological models in older clinical populations⁵¹. At the same time, this level of multicollinearity may weaken the validity of individual regression estimates and should be taken into account when interpreting the results. Accordingly, regression models should be interpreted as showing the contribution of the ACT Test within a multivariable cognitive context, rather than estimating independent effects of individual predictors.

To further evaluate diagnostic classification performance, ROC analyses were conducted. These analyses supported the diagnostic value of the ACT Test in distinguishing AD from healthy aging. The raw ROC curve showed strong discrimination ($AUC = 0.89$), and the Youden-derived cut-off of ≤ 39 yielded high sensitivity and specificity.

Beyond overall discrimination, the clinical utility of the ACT Test depends on how cut-off values are applied to different diagnostic decision-making contexts. When the primary aim is to minimize missed AD cases, a high-sensitivity cut-off (ACT Test ≤ 41 ; $Se \geq .95$) is more appropriate, as it provides high negative predictive value and supports rule-out decisions during initial clinical evaluation or referral. In contrast, when the goal is to strengthen diagnostic confidence and reduce false-positive classifications, high- or very high-specificity cut-offs (ACT Test ≤ 33 ; $Sp \geq .90$, or ACT Test ≤ 27 ; $Sp \geq .95$) should be preferable. These thresholds are better suited for rule-in decisions in patients with high clinical suspicion or when ACT Test findings are interpreted alongside other neuropsychological measures.

The Youden-derived cut-off (ACT Test ≤ 39) represents a balanced threshold and is most appropriate when no single clinical priority dominates and an overall measure of discrimination is desired. Accordingly, ACT Test cut-off values

should be selected based on clinical context rather than applied as fixed diagnostic rules.

In addition to cutoff-based decision strategies, demographic factors such as age play a critical role in diagnostic modeling. Adjusting for age improved diagnostic accuracy (AUC = 0.95), and the DeLong test confirmed this improvement was statistically significant ($z = 2.48, p < .05$). However, because statistical adjustment cannot fully remove the broad and interrelated effects of age on cognitive performance⁴⁷, the increased accuracy of the age-adjusted model should be interpreted with caution.

Specifically, the age-adjusted ROC reflects the discriminative performance of a combined logistic model including both ACT Test scores and age, rather than the diagnostic accuracy of the ACT Test alone. Thus, the higher AUC indicates that age contributes additional predictive value to the model, rather than demonstrating complete robustness of ACT Test performance to age-related influences.

ROC analyses were limited to comparisons between AD and healthy controls. The MCI group was excluded because it did not show significant differentiation from the control group in regression analyses ($p > .05$), which would likely have reduced the statistical reliability of a separate ROC model. Therefore, the diagnostic accuracy results reported here specifically reflect the ACT Test's discriminative validity for AD.

Within a multi-domain assessment framework, the ACT Test functions best as a complementary measure that provides specific information about interference-resistant verbal WM, rather than as a standalone screening tool. From a practical perspective, the ACT Test is a time-efficient measure and requires only standard neuropsychological administration skills. This makes it feasible for routine clinical use without extensive additional clinician training. When evaluated using Schulman's criteria⁵² for effective screening tests (brief administration, ease of scoring, minimal influence of language and education, and broad patient tolerability), the ACT Test is better suited as a diagnostic aid than as a general screening measure. Accordingly, its clinical utility is greatest when used to clarify the nature and extent of cognitive dysfunction by assessing WM as part of a comprehensive neuropsychological test battery, rather than for identifying early-stage impairment alone.

This study has several methodological limitations that should be considered when interpreting the findings. First, heterogeneity within the MCI group may have reduced sensitivity to subtle group differences. Variation in both memory impairment severity and subtype composition (amnesic vs. non-amnesic) is common at prodromal stages, where cognitive decline is less consistent. Such variability may have limited discrimination between MCI and healthy controls and increased the risk of sampling-related bias. As a result, generalizability to MCI populations with more clearly defined amnesic profiles may be constrained. Second, the recruitment strategy relied on individuals already engaged in specialty clinical services. This approach may have introduced selection bias by oversampling patients with greater caregiver support, higher health-seeking behavior, or more prominent cognitive concerns. Such factors may influence test performance and limit the applicability of the findings to routine or community-based clinical settings. Third, although age was statistically controlled, residual age-related influences likely remained. Aging affects multiple cognitive domains in a broad and overlapping manner, and a substantial proportion of age-related variance is shared across neuropsychological measures. As a result, statistical adjustment cannot fully isolate disease-specific impairment. Some observed group differences may therefore partly reflect normal age-related cognitive changes rather than pathology alone. Future studies should involve larger, more demographically balanced samples to more accurately isolate pathology-specific effects.

Measurement-related considerations also require attention. Although the ACT Test has been culturally adapted for healthy young adults in Turkish samples¹² and later refined with detailed administration and scoring guidelines¹⁴, its psychometric properties in healthy older adults have not been systematically validated. The lack of age-specific adaptation studies limits confidence in assuming equivalence across older populations. In addition, the cross-sectional design restricts conclusions about the temporal sensitivity. Cross-sectional data cannot determine whether the test captures longitudinal change or predicts progression from MCI to AD. Given that interference-resistant WM deficits may emerge later in the disease course, longitudinal studies are needed to determine whether ACT Test performance declines systematically with disease progression. Finally, the inability of the ACT Test to distinguish MCI from healthy controls represents a limitation of

the present data but also highlights directions for refinement. Sensitivity to prodromal deficits may be improved by increasing task difficulty, strengthening interference demands, or modifying delay intervals to increase cognitive load. Although the ACT Test may eventually be suitable for primary care or telehealth-based cognitive assessment, such applications should proceed cautiously and only after its clinical utility is confirmed through larger, longitudinal, and methodologically rigorous studies.

In summary, the present findings show that the ACT Test is a clinically useful measure of interference-resistant verbal WM and broader cognitive dysfunction, with strong validity (known-groups, convergent, and predictive) for AD but limited sensitivity to the subtler deficits seen in MCI. Its performance depends on multiple cognitive domains, including episodic memory, processing speed, and executive control, and is affected by demographic factors such as age. Although the ACT Test may serve as a diagnostic aid within a comprehensive neuropsychological test battery, its standalone value for early detection is limited. Future research with larger, demographically balanced, and longitudinal cohorts is needed to clarify its sensitivity to prodromal decline, refine task parameters, and assess its suitability for broader clinical use.

Author Contributions: Concept/Design : ACD, BCT, EST; Data acquisition: ACD; Data analysis and interpretation: ACD; Drafting manuscript: ACD, BCT; Critical revision of manuscript: ACD, BCT, EST; Final approval and accountability: ACD, BCT, EST; Technical or material support: ACD, BCT, EST; Supervision: ACD, BCT, EST; Securing funding (if available): n/a.

Ethical Approval: Ethical approval was obtained from the Hacettepe University Health Sciences Research Ethics Committee with decision number 2025/08-43 dated 08.04.2025.

Peer-review: Externally peer-reviewed.

Conflict of Interest: Authors declared no conflict of interest.

Financial Disclosure: Authors declared no financial support

REFERENCES

1. Madnani RS. Alzheimer's disease: A mini-review for the clinician. *Front Neurol.* 2023;14:1178588.
2. Rasmussen J, Langerman H. Alzheimer's disease - why we need early diagnosis. *Degener Neurol Neuromuscul Dis.* 2019;9:123-30.
3. Ding H, Wang B, Hamel AP, Melkonyan M, Ang TFA et al. Prediction of progression from mild cognitive impairment to Alzheimer's disease with longitudinal and multimodal data. *Front Dement.* 2023;2:1271680.
4. Öksüz N, Ghouri R, Taşdelen B, Uludüz D, Özge A. Mild cognitive impairment progression and Alzheimer's disease risk: A comprehensive analysis of 3553 cases over 203 months. *J Clin Med.* 2024;13:518.
5. Daou L, El Alayli A, Constantinos F, Dib G, Barakat M. Can neurocognitive assessment be a lower-cost substitute for biomarkers in predicting progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD)? A narrative review. *Biomark Neuropsychiatry.* 2023;9:100069.
6. Brown J. Some tests of the decay theory of immediate memory. *Q J Exp Psychol.* 1958;10:12-21.
7. Peterson LR, Peterson MJ. Short-term retention of individual verbal items. *J Exp Psychol.* 1959;58:193-8.
8. Stuss DT, Stethem LL, Poirier CA. Comparison of three tests of attention and rapid information processing across six age groups. *Clin Neuropsychol.* 1987;1:139-52.
9. Stuss DT, Stethem LL, Pelchat G. Three tests of attention and rapid information processing: An extension. *Clin Neuropsychol.* 1988;2:246-50.
10. Bherer L, Belleville S, Peretz I. Education, age, and the Brown-Peterson technique. *Dev Neuropsychol.* 2001;19:237-51.
11. Shura RD, Rowland JA, Miskey HM. Auditory consonant trigrams: A psychometric update. *Arch Clin Neuropsychol.* 2016;31:47-57.
12. Anil AE, Kivircik BB, Batur S, Kabakci E, Kitis A, Guven E et al. The Turkish version of the Auditory Consonant Trigram Test as a measure of working memory: A normative study. *Clin Neuropsychol.* 2003;17:159-69.
13. Callahan BL, Belleville S, Ferland G, Potvin O, Tremblay MP, Hudon C et al. Normative data for a computer-assisted version of the auditory three-consonant Brown-Peterson paradigm in the elderly French-Quebec population. *Clin Neuropsychol.* 2014;28:317-32.
14. Kademli M, Ertan-Kaya Ö, Salman F, Cangöz-Tavat B, Baran Z. The Auditory Consonant Trigram (ACT) Test: A norm updating study for university students. *Appl Neuropsychol Adult.* 2024;31:10-8.
15. Stuss DT, Stethem LL, Hugenholtz H, Richard MT. Traumatic brain injury: A comparison of three clinical tests, and analysis of recovery. *Clin Neuropsychol.* 1989;3:145-56.
16. Dannenbaum SE, Parkinson SR, Inman VW. Short-term forgetting: Comparisons between patients with dementia of the Alzheimer type, depressed, and normal elderly. *Cogn Neuropsychol.* 1988;5:213-33.
17. Morris RG. Short-term forgetting in senile dementia of the Alzheimer's type. *Cogn Neuropsychol.* 1986;3:77-97.
18. Ozakbas S, Ormeci B, Akdede BBK, Alptekin K, Idiman E. Utilization of the Auditory Consonant Trigram Test to screen for cognitive impairment in patients with multiple sclerosis: Comparison with the Paced Auditory Serial Addition Test. *Mult Scler J.* 2004;10:686-89.
19. Dige N, Wik G. Adult attention deficit hyperactivity disorder identified by neuropsychological testing. *Int J Neurosci.* 2005;115:169-83.

20. Baddeley AD, Hitch GJ. Working memory. In *The Psychology of Learning And Motivation*. Vol. 8 (Ed G Bower):47-89. New York, Academic Press, 1974.
21. Baddeley A. Working memory: Theories, models, and controversies. *Annu Rev Psychol*. 2012;63:1-29.
22. Oberauer K, Lewandowsky S. Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychol Rev*. 2008;115:544-76.
23. Mertens V, Gagnon M, Coulombe D, Messier C. Exploratory factor analysis of neuropsychological tests and their relationship to the Brown-Peterson task. *Arch Clin Neuropsychol*. 2006;21:733-9.
24. Aurtenetxe S, García-Pacios J, del Río D, López ME, Pineda-Pardo JA, Marcos A et al. Interference impacts working memory in mild cognitive impairment. *Front Neurosci*. 2016;10:443.
25. Hung SM, Wu DA, Shimojo S, Arakaki X. Stronger implicit interference in cognitively healthy older participants with higher risk of Alzheimer's disease. *Alzheimers Dement*. 2022;14:e12340.
26. Baddeley AD, Baddeley HA, Bucks RS, Wilcock GK. Attentional control in Alzheimer's disease. *Brain*. 2001;124:1492-508.
27. Bao X, Li W, Liu Y, Li X, Yue L, Xiao S. Impairment of delayed recall as a predictor of amnesic mild cognitive impairment development in normal older adults: A 7-year longitudinal cohort study in Shanghai. *BMC Psychiatry*. 2023;23:892.
28. Baddeley AD, Bressi S, Della Sala S, Logie R, Spinnler H. The decline of working memory in Alzheimer's disease: A longitudinal study. *Brain*. 1991;114:2521-42.
29. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149-60.
30. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup. *Alzheimer's Dement*. 2011;7:263-69.
31. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D et al. Practice guideline update summary: Mild cognitive impairment-Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Alzheimer's Dement*. 2018;14:280-92.
32. Güngen C, Ertan T, Eker E, Yaşar R. Standardize Mini Mental Test'in Türk toplumunda hafif demans tanısında geçerlik ve güvenilirliği. *Türk Psikiyatri Derg*. 2002;13:273-81.
33. Saka E, Mihci E, Topcuoglu MA, Balkan S. Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Arch Clin Neuropsychol*. 2006;21:745-51.
34. Ertan T, Eker E. Reliability, validity, and factor structure of the Geriatric Depression Scale in Turkish elderly: Are there different factor structures for different cultures? *Int Psychogeriatr*. 2000;12:163-72.
35. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
36. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-98.
37. Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. *Neurology*. 1988;38:900.
38. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills*. 1958;8:271-76.
39. Cangöz B, Karakoç E, Selekler K. Trail Making Test: Normative data for Turkish elderly population by age, sex and education. *J Neurol Sci*. 2009;283:73-8.
40. Cangöz B, Demirci S, Uluç S. Trail Making Test: Predictive validity study on Turkish patients with Alzheimer dementia. *Türk Geriatri Derg*. 2013;16:69-76.
41. Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders*. Philadelphia, Lea & Febiger, 1983.
42. Cangöz B, Karakoç E, Selekler K. The norm determination and validity-reliability studies of Clock Drawing Test on Turkish adults and elderly (ages 50 and over). *Türk Geriatri Derg*. 2006;9:136-42.
43. Karakaş S, Erdoğan E, Soysal Ş, Ulusoy T, Yüceyurt-Ulusoy İ, Alkan S. Stroop Test TBAG Form: Standardisation for Turkish culture, reliability and validity. *Türk J Clin Psychiatry*. 1999;2:75-88.
44. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatr Res*. 1983;17:37-49.
45. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-35.
46. Lezak MD, Howieson DB, Bigler ED, Tranel D. *Neuropsychological Assessment*. 5th ed. New York, Oxford University Press, 2012.
47. Salthouse TA. What needs to be explained to account for age-related effects on multiple cognitive variables? *Psychol Sci Public Interest*. 2010;11:65-87.
48. Belleville S, Lepage É, Hudon C, Bherer L, Gauthier S. The nature of working memory deficits in mild cognitive impairment and Alzheimer's disease. *Brain Cogn*. 2007;64:162-71.
49. Chang YL, Yang HT, Lin CP, Chiang HL. The independent and overlapping contributions of cognitive reserve and brain integrity to cognitive

- function in mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis.* 2015;47:15-26.
50. Jak AJ, Bondi MW, Delis DC, Salmon DP. Executive dysfunction in mild cognitive impairment. *Biol Psychiatry.* 2009;65:66-73.
 51. Salthouse TA. What cognitive abilities are involved in trail-making performance? *Intelligence.* 2011;39:222-32.
 52. Shulman KI. Clock-drawing: Is it the ideal cognitive screening test? *Int J Geriatr Psychiatry.* 2000;15:548-561.