UNIVERSITY OF PADOVA

Department of General Psychology

Bachelor's Degree Course in Psychological Science

Final dissertation

The Turkish Version of Auto-GEMS: Preliminary Data and Psychometric Properties

Supervisor:

Professor Dr. Mario Bonato

Co-supervisor:

Vincenzo Livoti

Candidate: Defne Naz Konuşmaz

Student ID number: 2121978

Academic Year 2025-2026

Abstract

Recent technological developments have brought an increase in computerized assessment and testing in many areas of scientific research and healthcare, and neuropsychology is no exception. Through the rise in computerized assessments and digital tests, remote testing is made possible for diverse populations that otherwise have little to no access to tests that monitor their cognitive functioning. Auto-GEMS is a self-administered web-based cognitive screening test that gives the possibility to quickly measure an individual's cognitive state along with their cognitive reserve, initially normed for the Italian population only. Auto-GEMS has been translated to several languages and many of them are currently in the phase of validation. In this study on the recently translated Turkish version of Auto-GEMS, data on a sample of 29 participants were collected with the aim of evaluating its psychometric properties. The first chapter introduces the concept of computerized assessment and digital testing in neuropsychology, explains the psychometric properties all online cognitive tests are expected to have, and names the objectives of the study. The second chapter demonstrates the methods and materials used in the study, the procedure and the statistical analyses. The third chapter presents the results obtained from the analyses, and the fourth chapter summarizes and discusses the findings by addressing limitations and suggestions for future research.

Keywords: digital tools, neuropsychology, web-based tests, cognitive screening, normative data

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Over the past years, numerous instruments have been designed to evaluate the global state and cognitive functioning of healthy populations and of neurological patients in research or healthcare settings. The rigorous methodology of cognitive test development allows researchers to investigate cognitive functioning through valid, reliable and standardized measures, ensuring the appropriate application of such tests and the interpretation of results. A transformation can be seen across all scientific disciplines, including neuropsychology, through the integration of digital technologies into practice. According to Parsons (2016), the different phases of development of neuropsychological assessment can be explained in parallel with surges in technological advancements, and the next generation of neuropsychological applications seems to be defined by increased Internet accessibility. Given the era we are in, the issue of digital cognitive testing warrants greater research attention so that the fields of use can be expanded.

This prospective study was designed to further develop and investigate the newly translated Turkish version of the web-based cognitive screening test Auto-GEMS by collecting data and conducting statistical analyses to demonstrate initial psychometric properties. Digital cognitive testing can help researchers gather large normative databases in a short amount of time, which are stored in centralized data repositories (Tuerk et al., 2023). However, in the field of neuropsychology, there seems to be a reluctance to integrate data science methods due to limited familiarity with data science and a lack of consensus regarding the types of data management tools to use (Singh & Germine, 2021). This study is especially beneficial for data sharing, considering there are very few online cognitive tests shared worldwide with accessible psychometric data (Charalambous et al., 2020). It is hoped that this study will contribute to the growing area of research on digital neuropsychological tools in various languages, their future adaptation and validation, and consequent widespread utilization.

1.1 Cognitive Screening Tests

The general purpose of cognitive screening tests is to make inferences about the likelihood of cognitive impairment from comparing the patient's score to reference norms (Cullen et al., 2007). Through this process, it is possible to obtain indices of various cognitive domains in a short amount of time, which is relevant for assessing cognitive status, and may identify a need to get further neuropsychological consultation. The availability of brief, remotely administered, web-based tests can further facilitate monitoring and detecting changes in cognitive functioning. In research settings, digital cognitive screening can aid in participant selection by confirming that all individuals meet certain criteria, consequently helping with obtaining a more accurate interpretation of research outcomes and increasing data quality. Additional benefits in research settings include establishing baselines for populations and collecting large-scale data for epidemiological studies. It is important to keep in mind that online tools seek to improve the ways that data is collected, and to interpret the data, a skilled professional is needed, nonetheless.

1.1.1 Auto-GEMS

Auto-Global Examination of Mental State (Auto-GEMS) is a recent, web-based, self-administered cognitive screening tool that gives the possibility to quickly measure an individual's cognitive state, also considering their cognitive reserve (Pucci et al., 2024). It has been originally normed for the Italian population only. Auto-GEMS is the computer-based and self-administered version of the paper-and-pencil-based test GEMS (Mondini et al., 2022) and its remotely administered counterpart Tele-GEMS (Montemurro et al., 2023).

With the development of Auto-GEMS, researchers have provided a way to make cognitive screening more accessible, as the test can be made available in many languages and can be completed independently, with or without supervision, on any computer, through a web browser, without special software, at any time and from any location. As of now, adaptations are being developed in English, Turkish, Spanish, Farsi and Russian. Auto-GEMS has two parallel versions called A and B, consists of 11 tasks that can be completed in around 10 minutes and provides a reliable evaluation of the global state of cognitive functioning, along with calculating examinees' cognitive reserve with a shortened version of the Cognitive Reserve Index questionnaire (Nucci et al., 2012) in relation to their demographic information.

For the original version in Italian, cut-off scores were established through a regression-based approach (Crawford & Garthwaite, 2006) that identified predictors of scores. In this case, potential predictors were named to be age, sex, education and cognitive reserve.

1.1.2 Cognitive Reserve

Cognitive reserve is the process by which the brain utilizes pre-existing cognitive processing approaches and compensatory strategies in an attempt to offset brain damage (Stern, 2017). The concept of cognitive reserve refers to the hypothesis developed to explain the discrepancy between the degree of brain damage or pathology and its clinical manifestations and variations in disease outcomes (Stern, 2009). A longitudinal study following a cohort aged 75-85 years at intake by Katzman et al. (1989) has revealed that some presumably non demented patients with normal cognitive functioning had advanced AD (Alzheimer's Disease), found in a postmortem examination. These findings, along with other observable individual differences in brain disease outcomes, illustrate that these differences may be attributed to the presence of a compensatory reserve.

Cognitive reserve acts as a buffer and allows patients with higher cognitive reserve to use cognitive processes flexibly to cope better with the consequences of brain pathology. The hypothesis suggests that people with higher cognitive ability, along with the factors and influences associated with it, have a lower risk of dementia than individuals with less cognitive ability (Whalley et al., 2004). Some factors implicated in the development of someone's cognitive reserve over the lifetime are education, occupation, intelligence and leisure activity. The Cognitive Reserve Index questionnaire was developed with the purpose of offering a measure of cognitive reserve by calculating the amount of CR individuals have gathered throughout their lifetimes (Nucci et al., 2012).

1.2 Computerized Assessment and Digital Testing in Neuropsychology

Recent technological developments suggest a shift towards the increase and sometimes the preference of computerized neuropsychological assessment and digital cognitive testing in both research and clinical settings. Neuropsychological assessment makes use of measurement tools and tests, which are recognized as "reliable" and "valid", and consists of the comprehensive and structured examination of different cognitive functions (Cipolotti & Warrington, 1995). As part of neuropsychological assessment, cognitive testing supports interpretation through standardized evaluation of global cognitive functioning and identification of cognitive domain deficits by examining mental processes such as memory, language, executive function, abstract reasoning, attention, and visuospatial skills (Watt et al., 2021).

Computerized assessment and digital testing devices in neuropsychology include computer-administered versions of existing tests and new tests particularly designed to be completed on a computer with general and specific applications, namely for cognitive screening or for sensitively detecting conditions impacting cognitive functioning (Bauer et al., 2012). There are many reasons for this shift: computerized tests are cost-effective, offer flexibility in administration and allow researchers and clinicians to access a larger and more diverse population demographic (Morrissey et al., 2024). Computerized batteries are efficient in administration as they can collect high frequencies of data while having a faster pace, in addition to being more precise in capturing responses in time-sensitive tasks and providing parallel versions with alternate sets of stimuli to minimize practice effects (Huynh et al., 2025). The built-in standardization afforded by the nature of computerized assessments increases reproducibility and reliability by default and reduces the possibility of administrator error in recording, scoring and interpreting participants' responses by automating the procedure (Miller & Barr, 2017). As a sub-type of computerized assessments, self-administered web-based tests can be seen as more adaptable, as they can be accessed on a personal computer in a familiar setting without the need to download specific software. One of the biggest advantages of web-based computerized assessments is offering services to people in remote areas, to people who don't have the privilege of getting routine consultations, and to populations less likely to seek cognitive evaluations, eliminating disparities in access to healthcare (Tsoy et al., 2021). Web-based computerized tests are usually readily available and translated to several languages, normed for various populations, further allowing neuropsychologists to reach a wider demographic and to make cross-cultural comparisons.

In contrast, the traditional, supervised paper-based neuropsychological assessment has limited application due to being time-consuming and labour intensive, consequently allowing less people to undergo cognitive assessment. Another limitation of traditional neuropsychological tests is that they have restricted ecological validity as they are conducted in artificially controlled environments such as a clinic or a laboratory. Cognitive performance is influenced by many factors, such as fatigue, mood and time of day, that can't be measured in a simulated and standardized setting (Harris et al., 2024). On the other hand, self-administered online tests aim to capture a more realistic picture of daily cognitive functioning as participants can conveniently take the test at any time at a location of their choice. In a study by Hansen et al. (2015), participants' feedback interviews revealed that they perceived the self-administered web-based battery as less difficult and felt more in control without a test administrator present compared with traditional tests, which might lead to more accurate evaluations without the constraint of test-taking and observer-related anxiety.

It is important to report that online tools also have a number of shortcomings which prevents neuropsychological testing to be completely dependent on digital technologies. Since they are usually administered in an unmonitored setting, no qualitative observation can take place, which is an integral part of neuropsychological assessments. The lack of supervision may also make participants more prone to distractions or may cause them to misinterpret the instructions (Backx et al., 2020; Feenstra et al., 2018). Managing technical difficulties that hinder test-taking might be challenging without supervision, and many older participants might require additional help navigating the computer. Technological issues, such as variations in computer hardware, software and in computer skills, are among the limitations of computerized assessments.

A study by Iverson et al. (2009) investigating the association between familiarity with computers and performance on computerized neurocognitive testing demonstrated that participants with "some" computer use had significantly lower psychomotor speed, reaction time, complex attention and cognitive flexibility domain scores and did worse on tasks requiring visual scanning and mouse or keyboard work compared to participants with "frequent" computer use. Contextual differences such as computer familiarity can negatively or positively affect performance and alter the scores obtained on unsupervised web-based tests, which should be taken into consideration to improve data quality, result interpretation and diagnostic accuracy.

Due to the advantages and disadvantages listed, online cognitive tests are currently advised to be used as a pre-screening tool in clinical settings and to screen participants for clinical trials in research settings (Morrissey et al., 2024).

1.3 Psychometric Properties of Online Cognitive Tests

The scores of neuropsychological tests are obtained from a person's performance compared with the performance of a healthy normative sample, clinical samples, one's expected level of performance according to their demographics or research participants who were given standardized and specific instructions (Bauer et al., 2012). Before interpreting test scores, it is crucial to ensure that a cognitive test shows a certain level of reliability and validity, and that the normative data is available and is representative of the population.

One of the most mentioned concerns regarding online computerized cognitive tests is that many don't have thoroughly examined and well-validated psychometric properties, especially for clinical populations. Since the issue of online cognitive tests is an emerging area, many tests seem to lack strong psychometric support for measures of validity and reliability (Van Patten, 2021). Validity refers to the extent to which an instrument measures what it is supposed to measure (Kimberlin & Winterstein, 2008), and reliability refers to the extent to which an instrument consistently has the same results on similar occasions (Heale & Twycross, 2015). During the test development phase, all psychological tests are required to meet specific criteria to determine their usability in various settings.

Establishing the validity and reliability of self-administered web-based cognitive tests is essential, especially because the examinees complete the test independently, often without supervision, and are subject to distinct limitations related to the test setting that may influence scores. Examining psychometric properties such as test-retest reliability, practice effects and normative change is necessary to ascertain if the change in scores is due to actual changes in cognition or due to confounding variables related to the testing process (Bergman et al., 2025). Extensive normative data on different populations should be collected, and databases should be made available to investigate cognitive changes across a variety of languages, cultures, age groups and education levels. Much more research attention should be given to investigating the psychometric rigor of digital neurocognitive tests, as there seems to be a discrepancy between the fast-paced development of digital software and empirical evidence backing the massive surge in test batteries (Van Patten, 2021).

1.3.1 Online Cognitive Tests and Demographic Diversity

Although online cognitive tests have the advantage of being administered in any place and reaching a wider demographic, multiple reviews report variability in methodology and psychometric properties across tests, along with a lack of diversity in the sample (Cubillos & Rienzo, 2023; Tsoy et al., 2021; Wild et al., 2008). These findings illustrate the need for further development, validation and evaluation of test properties of computerized cognitive tests in different languages and demographic samples to broaden and facilitate their implementation. By leveraging the Internet, researchers can now reach many different populations across the world, without geographical constraints. This advantage holds promise, especially since there is a disparity in the amount of available cognitive tests and neuropsychological development between nations (Fernandez, 2019).

In consideration of such gaps, researchers should be careful in integrating modern tools into their routines. Future tests should be cross-culturally validated, appropriate for all types of educational backgrounds, digitally inclusive for developing regions, affordable and economically accessible for everyone (Fernandez, 2019). Digital literacy levels, strategies and frameworks differ amongst countries; therefore, a cross-cultural approach is necessary to investigate inequalities and improve application. Demographic variables have also been shown to directly influence an individual's access to digital technology and digital literacy skills: for Turkish adults, gender, age, and educational status seem to be predictors of digital literacy (Cicikler et al., 2025).

Regardless of the medium, neuropsychological tests appear to be subject to cross-cultural biases, such as construct, method and item bias, and language influences (Fernández & Abe, 2018). For bilingual individuals, performance in neuropsychological tests has been shown to be dependent on the language version of the test. In a study by López et al. (2016), bilingual participants differed in their performance on a digit-span task, depending on whether they were tested in English or Spanish, hypothetically because of variations in cognitive strategies used by diverse language groups. A review by Celik et al. (2022) also reported that, for bilingual older adults, neuropsychological test performance was seemingly influenced by language proficiency, age of acquisition of the second language, language of test administration and the use of culturally/linguistically appropriate tools. Across the tests critically examined, reports of participants' language proficiency were very few, and many tests overlooked the potential impact of cultural and linguistic variables on test performance. Researchers are advised to be informed of the cultural and linguistic factors that may influence online cognitive test performance and consider the diversity of the sample during the test development phase.

The aim of this study is to conduct a preliminary assessment of the psychometric properties of the recently translated Turkish version of the self-administered cognitive screening tool Auto-GEMS by demonstrating data, providing an initial measurement of its reliability and investigating correlations between two language versions (Turkish and English). By employing quantitative and qualitative methods, the study attempts to discuss the current state of the Turkish Auto-GEMS in a relatively small sample, to contribute to the ongoing development of the language version.

CHAPTER 2: METHODS

2.1 Participants

A sample of 29 Turkish-speaking participants was recruited (23 Females, 6 Males) using convenience sampling. Participation was completely voluntary, and all participants were contacted directly through email with a personalized ID, instructions and a single access link that redirected them to an anamnestic questionnaire to ensure that they fit the inclusion criteria. All participants were required to be adults over the age of 18 years old, have a certain degree of computer knowledge, and not have any psychiatric, neurological, or any other disorder that may impact the test results. The sample was controlled before analysis, and no exclusions were made according to the criteria. Before proceeding with the test, participants were asked to read and agree to an informed consent form, and they were allowed to withdraw from the test at any moment.

2.2 Procedure

The current study had a test-retest design, which is reported in figure 2.1. A total of 29 participants were recruited. For the first phase of the study, 22 participants were tested with the Turkish version of Auto-GEMS, and 7 participants were tested with the English version of Auto-GEMS, either in presence or remotely. Only a sub-sample of 6 participants were tested in presence and before the Auto-GEMS, the Turkish MoCA was administered and scored. The second phase of the study took place one month after and was conducted remotely for all participants, using the alternate language version: a sub-sample of 9 participants, who were initially tested with the Turkish version, were tested remotely with the English version. Another sub-sample of 7 participants, who were initially tested with the English version, were tested with the Turkish version. Across both phases, 29 participant responses for the Turkish Auto-GEMS and 16 participant responses for the English Auto-GEMS were collected. Different from the Turkish version, the English Auto-GEMS included an additional questionnaire assessing language level and proficiency after the anamnestic questionnaire and before the test. A link containing the relevant language version of the test was sent to the participants via email, along with a personalized ID to maintain anonymity. For the sub-sample ($N=6$) tested in person, the Turkish MoCA translated by Selekler et al. (2010) was administered beforehand. Participants were then required to carefully read the email sent to them, which mentioned the aims of the study along with the instructions detailing that the test must be taken in a quiet environment without distractions, must be done independently, must be accessed using only Chrome or Firefox browsers on a computer and must be completed in a single session without interruptions. After reading the instructions, participants were asked to open the link, which redirected them to the anamnestic questionnaire, consisting of questions about their medical history, and the Memory Assessment Clinics Questionnaire (MAC-Q).

After completing the questionnaires on the platform Qualtrics, participants were randomly assigned to one of the two versions of the test (Auto-GEMS A or Auto-GEMS B) through the platform JATOS (Just Another Tool for Online Studies). Before starting the behavioural task, participants agreed to the informed consent form, re-entered their IDs and gave their demographics. A shortened version of the Cognitive Reserve Index questionnaire (CRIq), which contained 6 items about the level of education, occupation, type and frequency of leisure activities, was presented. Following the CRIq, a brief audio check was performed, and the Auto-GEMS (A or B) started. After successful completion of the test, participants filled out a questionnaire on their testing experience and were encouraged to give feedback. All data from the Auto-GEMS test were collected on JATOS and automatically sent back to the server of the Department of General Psychology at the University of Padova. Data files in .csv format were also saved on the participant's computer, containing their global score, scores in each task, questionnaire responses and CRIq scores.

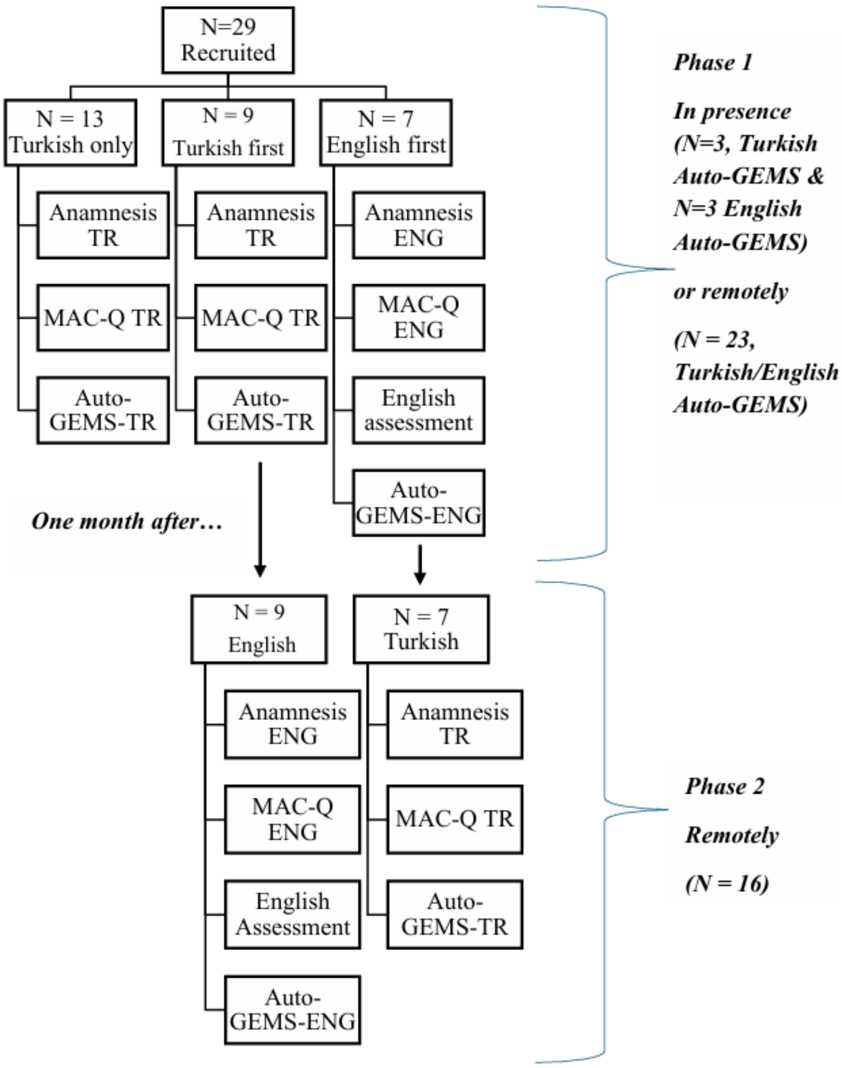

Figure 2.1. Data Collection Design

2.3 Materials

For this study, the following instruments were used:

1. MoCA-TR: The Turkish version translated and adapted by Selekler et al. (2010) was administered in person. MoCA is administered in 10 minutes, is scored out of a maximum of 30 points and measures eight cognitive domains: short-term memory; visuospatial abilities; executive function; attention, concentration, and working memory; language; temporal and spatial orientation (Nasreddine et al., 2005). To explain briefly, the tasks are short term memory recall (5 points), clock-drawing (3 points), three-dimensional cube copy (1 point), TMT-B (1

point), phonemic fluency (1 point), verbal abstraction (2 points), sustained attention (1 point), serial subtraction (3 points), digits forward and backward (1 point each), naming task (3 points), complete repetition of sentences (2 points), orientation of time and place (6 points). The cut-off score of < 21 points for the Turkish version was taken as a reference.

2. Medical History Questionnaire: Collects information about current or past health issues that might affect cognitive functioning. Includes questions on disabilities, neurological complications such as seizures, head trauma and epilepsy, psychiatric conditions, substance use, sleep disturbances and medication intake.

3. MAC-Q: Contains 6 items that rate memory performance on a 5-point scale from "much worse now" to "much better now". Aims to assess the subjective perception of age-related memory decline by comparing current performance to that of 10 years ago.

4. English language proficiency assessment: A questionnaire asking participants for details on how well and how often they use the English language was presented to assess their language proficiency. The assessment included items on how often they use English, how long they've been using it, in what settings they use it, how well they can speak and understand English even if the subject matter is complex, how confident they are in using it, and if applicable, the level of the language certification they've obtained. Participants were given a score for each item on the questionnaire, which was then summed up to make the total score.

5. CRIq: 6 items adapted from the original 20-item CRIq were utilized. The full version of the CRIq includes a demographic questionnaire followed by 20 items measuring the scales CRI-Education (years of education and other training courses), CRI-WorkingActivity (adulthood professions with the cognitive level of the job) and CRI-LeisureTime (years and frequency of stimulating activities during leisure time). At the end, the raw scores obtained from the scales are correlated with age by the number of years the examinee has been doing the activity.

6. Auto-GEMS-A and Auto-GEMS-B: The recently translated Turkish version and the English version of Auto-GEMS were used. The parallel versions A and B share the same structure but differ in item content. The items consist of:

- *Orientation:* Participants are asked three questions on temporal and spatial orientation, such as the year they're in.

- *Immediate memory:* Participants are briefly presented with six words and are later asked to type in the words they can remember.

- *Months backwards:* Participants are asked to type in reverse order the months of the year by skipping one month at a time, starting from October.

- *Puzzle:* Participants are instructed to match four pieces of a figure through a drag and drop system.

- *Spatial representation:* Participants are asked to visualize and report on which side (right or left) the two hands of an imaginary clock are positioned according to a given time.

- *Delayed memory*: Participants are asked to recall and type in the same six words presented previously in the immediate memory task.

- *Naming:* Participants are shown four objects and are asked to type in their names.

- *Comprehension:* Participants are ordered to "Press the letter 'A' twice after pressing the letter 'B' once".

- *TMT-A:* Participants are requested to click on numbers from 1 to 14 in ascending order as quickly as they can. Response accuracy and processing speed in seconds are calculated and considered when weighing the tasks.

- *TMT-B:* Participants are asked to click on and connect numbers from 1 to 7 in ascending order with letters from “A” to “G” in alphabetical order by alternating between them. Response accuracy and processing speed in seconds are calculated and considered when weighing the tasks.

- *Metaphor:* Participants are asked to choose between three possible interpretations of a metaphorical sentence.

For each correct response and execution, one point is given. Each task results in a raw score, which is then transformed so that the tasks weigh the same, on a range from 0 to 9. The global score is calculated automatically by summing the weighted scores.

7. End of Test Feedback Questionnaire: Participants were asked to pick the option(s) that best described their testing experience.

2.4 Data Analysis

Data from the questionnaires and the behavioural task were collected on Qualtrics and JATOS, respectively, and were automatically uploaded to the server connected to the University of Padova. The data used for the following statistical analyses were obtained from the server in the form of .csv files and were imported to RStudio (version 2026.04; Posit team, 2026) to perform the data merging, cleaning and filtering according to the IDs. During this process, the mismatched IDs of a few participants were also corrected for. The cleaned and filtered dataset was then imported to the open-source platform JASP (version 0.97.1; JASP Team, 2026) [Computer software] to be analysed. In the following section, descriptive statistics of the sample are reported for demographic variables, Cognitive Reserve scores, and the scores obtained from the Turkish Auto-GEMS. Additionally, multiple regression analyses are made to assess the relationship of age, education and Cognitive Reserve with the Turkish Auto-GEMS. The internal consistency across items and the correlations of each item with the global score are calculated. Correlations between the scores of the Turkish and English versions are demonstrated. The correlation of English proficiency with the global score of the English Auto-GEMS is investigated. Potential practice effects are examined through a paired sample t-test between the two language versions according to the order of administration. Feedback received from the participants on their testing experiences is reported.

CHAPTER 3: RESULTS

Descriptive statistics of the demographics of the sample revealed that the mean age of the sample was 40.76 years (*SD* = 14.97; ranging from 20 to 67), with a slightly left skewed (skewness = -0.10) distribution, indicating that the distribution was approximately symmetric with a small tendency for older participants. The mean education was 16.21 years (*SD* = 1.95; ranging from 13 to 19), with a moderately left skewed (skewness = -0.53) distribution, suggesting that most participants had a high level of education. In the sample of the study, the mean cognitive reserve index score measured in the Turkish Auto-GEMS was 126.5 (*SD* = 16.34; ranging from 107 to 165), and the distribution was asymmetrical and positively skewed (skewness = 1.18). Additional descriptive statistics of the Turkish Auto-GEMS are presented in Table 3.1.

	Median	Mean (arithmetic)	Std. Deviation	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Minimum	Maximum
Age	45.00	40.76	14.97	−0.102	0.434	−1.504	0.845	20.00	67.00
Education	16.00	16.21	1.953	−0.528	0.434	−0.764	0.845	13.00	19.00
CRI	120.0	126.5	16.34	1.179	0.434	0.407	0.845	107.0	165.0
Auto-GEMS-TR Global Score	80.00	79.45	10.05	−0.887	0.434	1.912	0.845	50.00	98.00
Orientation	3.000	2.621	0.494	−0.525	0.434	−1.858	0.845	2.000	3.000
Immediate Memory	5.000	4.655	1.078	−0.524	0.434	−0.183	0.845	2.000	6.000
Spatial Representation	4.000	3.793	0.559	−2.698	0.434	6.381	0.845	2.000	4.000
Delayed Memory	3.000	3.241	1.480	−0.021	0.434	−0.736	0.845	1.000	6.000
Puzzle	3.000	3.000	0.000	NaN[a]	0.434	NaN[a]	0.845	3.000	3.000
Naming	4.000	3.448	0.827	−1.449	0.434	1.471	0.845	1.000	4.000
Comprehension	3.000	2.310	1.137	−1.293	0.434	0.021	0.845	0.000	3.000
TMT-A Accuracy	14.00	13.17	3.140	−3.768	0.434	13.58	0.845	0.000	14.00
TMT-A Processing Speed	13.67	15.86	7.196	2.453	0.434	8.145	0.845	8.037	44.37
TMT-B Accuracy	14.00	11.24	4.033	−1.182	0.434	0.266	0.845	1.000	14.00
TMT-B Processing Speed	18.49	19.91	8.627	1.202	0.434	3.039	0.845	3.868	48.07
Metaphor	1.000	0.966	0.186	−5.385	0.434	29.00	0.845	0.000	1.000
Months Backwards	1.000	1.552	1.617	0.745	0.434	−1.143	0.845	0.000	4.000

[a] All values are identical

Table 3.1. Descriptive Statistics for Demographic Variables, Turkish CRI Scores, and Auto-GEMS-TR Scores

The mean global score for the Turkish Auto-GEMS was 79.45 (*SD* = 10.05; ranging from 50 to 98). The distribution of total scores was moderately left skewed (skewness = -0.89) and thin-tailed (kurtosis = 1.91), demonstrating that higher scores were more commonly present and outliers were not frequent. Figure 3.1 presents a boxplot of the distribution of scores, including the identified outliers. The outliers were examined and were found to be representative of real results and consequently were included in the data analysis.

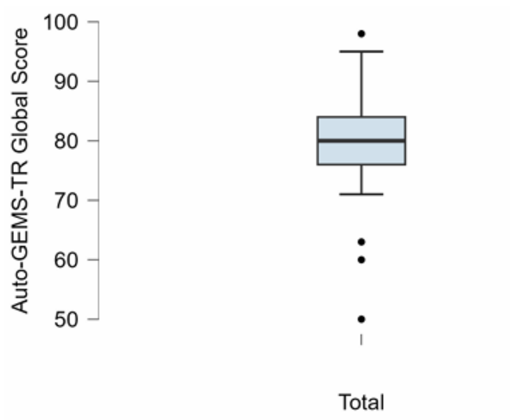

Figure 3.1. The distribution of Turkish Auto-GEMS Global Scores

A multiple linear regression model was carried out to examine the relationship between Age, Education, Cognitive Reserve with Global Turkish Auto-GEMS scores, as they were named among the strongest predictors.

Model	R	R²	Adjusted R²	RMSE	R² Change	df1	df2	p
M_0	0.000	0.000	0.000	10.05	0.000	0	28	
M_1	0.434	0.188	0.091	9.580	0.188	3	25	.150

Note. M_1 includes CRI, Education, Age

Model		Sum of Squares	df	Mean Square	F	p
M_1	Regression	532.8	3	177.6	1.935	.150
	Residual	2294	25	91.78		
	Total	2827	28			

Note. M_1 includes CRI, Education, Age

Model		Unstandardized	Standard Error	Standardized	t	p
M_0	(Intercept)	79.45	1.866		42.58	< .001
M_1	(Intercept)	114.3	22.15		5.161	< .001
	CRI	−0.027	0.127	−0.044	−0.213	.833
	Education	−1.756	1.204	−0.341	−1.458	.157
	Age	−0.073	0.174	−0.109	−0.421	.677

Table 3.2. Multiple Linear Regression Model that tested the effect of Age, Education, and Cognitive Reserve on Turkish Auto-GEMS Scores

According to the preliminary analysis, age was not a significant predictor of scores (B = -0.07, SE = 0.17, β = -.11, t = -0.42, p = .677). Similarly to Age, Cognitive Reserve was not a significant predictor of scores (B = -0.03, SE = 0.13, β = -.04, t = -0.21, p = .833). Education was a relatively stronger predictor of scores but still was not statistically significant (B = -1.76, SE = 1.20, β = -.34, t = -1.46, p = .157). Overall, the regression model did not significantly predict Turkish Auto-GEMS global scores, (F = 1.94, p = .150) and explained only the 19% of variance in the scores (R^2 = .19, Adjusted R^2 = .09). These results are illustrated in Table 3.2.

The internal consistency of the Turkish Auto-GEMS was measured by Cronbach's α on all items, and the results revealed poor internal consistency, α = .49 (see Table 3.3).

			95% CI	
Coefficient	Estimate	Std. Error	Lower	Upper
Cronbach's α	0.485	0.087	0.314	0.656

Table 3.3. Internal consistency across items

The items Immediate Memory, r = .65, Naming, r = .48, Months Backwards, r = .40, Delayed Memory, r = .68, TMT-B Accuracy, r = .43, Metaphor, r = .56, Comprehension, r = .59, were significantly correlated with the global Auto-GEMS score. Overall, there was a medium-to-large positive association between the listed items and the total score. The correlation of the item Orientation, r = .34, showed a moderate effect size, and the correlation of the item Spatial Representation, r = .16, showed a small effect size with the global score but these did not reach statistical significance. The processing speed recorded for the TMT-A task also showed a weak-to-moderate negative association with the global score, r = -.49. Correlations between test items and Turkish Auto-GEMS scores, along with the p-values and confidence intervals are illustrated in Table 3.4.

			Pearson's r	p	Lower 95% CI	Upper 95% CI
Auto-GEMS-TR Global Score	-	Orientation	0.338	.073	−0.033	0.627
Auto-GEMS-TR Global Score	-	Immediate Memory	0.648***	< .001	0.369	0.820
Auto-GEMS-TR Global Score	-	Months Backwards	0.402*	.031	0.042	0.670
Auto-GEMS-TR Global Score	-	Spatial Representation	0.157	.416	−0.222	0.495
Auto-GEMS-TR Global Score	-	Delayed Memory	0.682***	< .001	0.420	0.839
Auto-GEMS-TR Global Score	-	Naming	0.478**	.009	0.135	0.718
Auto-GEMS-TR Global Score	-	Comprehension	0.591***	< .001	0.286	0.787
Auto-GEMS-TR Global Score	-	TMT-A Accuracy	0.082	.671	−0.293	0.436
Auto-GEMS-TR Global Score	-	TMT-B Accuracy	0.434*	.019	0.080	0.690
Auto-GEMS-TR Global Score	-	TMT-A Processing Speed	−0.493**	.007	−0.728	−0.154
Auto-GEMS-TR Global Score	-	TMT-B Processing Speed	−0.205	.287	−0.531	0.175
Auto-GEMS-TR Global Score	-	Metaphor	0.564**	.001	0.248	0.771

Table 3.4. Correlations between the global score and the test items for the Turkish Auto-GEMS

For the sub-sample *(N=16)* tested also with the English Auto-GEMS, the global scores obtained from the Turkish and English versions showed a moderate positive correlation, r = .48, p = .058 (See Table 3.5), but this relationship did not reach statistical significance. There was a strong statistically significant positive correlation between language proficiency and the global scores of the English Auto-GEMS, r = .73, p = .001. (See Table 3.6 and Figure 3.2).

			Pearson's r	p
Auto-GEMS-ENG Global Scores	-	Auto-GEMS-TR Global Scores	0.483	.058

Table 3.5. Correlation between the global scores of both versions (Auto-GEMS-ENG & Auto-GEMS-TR)

			Pearson's r	p
Auto-GEMS-ENG Global Scores	-	English Proficiency	0.725	.001

Table 3.6 Correlation between the global score of Auto-GEMS-ENG and English Proficiency

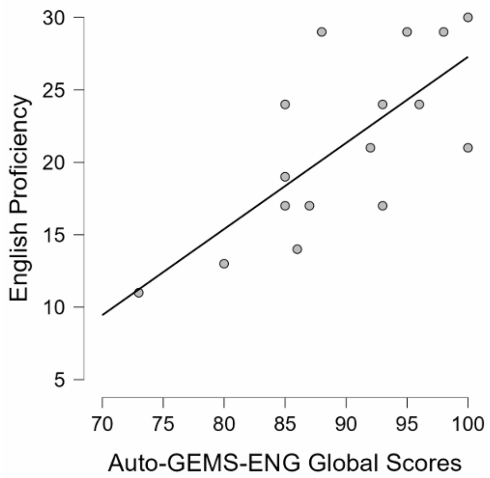

Figure 3.2. Correlation between Auto-GEMS-ENG Global Scores and English Proficiency

Practice effects were explored through a paired sample t-test for participants who took the English version first and for participants who took the Turkish version first, to see if performance changed over repeated sessions. For the participants who took the English Auto-GEMS first and the Turkish Auto-GEMS second, the global score decreased from the first administration (M = 91.71, SD = 7.46) to the second administration (M = 83.57 SD = 7.04), $t(6)$ = -2.45, p = 0.05, d = -.93, however this decrease was not statistically significant. The scores for the Months Backwards item decreased significantly from English (M = 4.43, SD = 0.54) to Turkish (M = 1.57, SD = 1.72), $t(6)$ = 4.26, p = 0.005, d = 1.61. Differences between the global score and the single task scores can be seen in Table 3.7, along with the means in Table 3.8.

								95% CI for Cohen's d	
Measure 1		Measure 2	t	df	p	Cohen's d	SE Cohen's d	Lower	Upper
T2: Auto-GEMS-TR	-	T1: Auto-GEMS-ENG	−2.448	6	.050	−0.925	0.548	−1.800	-1.784×10^{-4}
Immediate Memory-ENG	-	Immediate Memory-TR	0.000	6	1.000	0.000	0.402	−0.741	0.741
Months Backwards-ENG	-	Months Backwards-TR	4.264	6	.005	1.612	0.789	0.430	2.745
Delayed Memory-ENG	-	Delayed Memory-TR	1.362	6	.222	0.515	0.556	−0.297	1.290
Naming - ENG	-	Naming-TR	−0.750	6	.482	−0.283	0.691	−1.030	0.485
Comprehension-ENG	-	Comprehension-TR	1.000	6	.356	0.378	8.245×10^{-9}	−0.406	1.134

Note. Student's t-test.

Table 3.7. Paired sample t-test: Auto-GEMS-ENG vs. Auto-GEMS-TR

	N	Mean	SD	SE	Coefficient of variation
T1: Auto-GEMS-ENG	7	91.71	7.455	2.818	0.081
T2: Auto-GEMS-TR	7	83.57	7.044	2.662	0.084
Immediate Memory-ENG	7	4.857	1.215	0.459	0.250
Immediate Memory-TR	7	4.857	0.900	0.340	0.185
Months Backwards-ENG	7	4.429	0.535	0.202	0.121
Months Backwards-TR	7	1.571	1.718	0.649	1.093
Delayed Memory-ENG	7	4.571	1.988	0.751	0.435
Delayed Memory-TR	7	3.286	1.604	0.606	0.488
Naming - ENG	7	3.143	0.900	0.340	0.286
Naming-TR	7	3.571	0.787	0.297	0.220
Comprehension-ENG	7	2.857	0.378	0.143	0.132
Comprehension-TR	7	2.714	0.756	0.286	0.279

Table 3.8. Mean scores for both conditions (Auto-GEMS-ENG vs. Auto-GEMS-TR)

As seen in Table 3.9 and 3.10, for the participants who took the Turkish Auto-GEMS first and the English Auto-GEMS second, the global score decreased significantly from the first administration (*M* =87.44, *SD* = 8.26) to the second administration (*M* = 78.89 *SD* = 8.10), *t*(8) = -2.78, *p* = .024, *d* = -.93. The scores for the item Months Backwards increased significantly from Turkish (*M* = 0.67, *SD* = 1.32) to English (*M* = 4.67, *SD* = 0.50), *t*(8) = 9.07, *p* < .001, *d* = 3.02. No practice effect was observed for both language retests, but rather, a decrease in global score was seen with large effect sizes.

								95% CI for Cohen's d	
Measure 1		Measure 2	t	df	p	Cohen's d	SE Cohen's d	Lower	Upper
T2: Auto-GEMS-ENG	-	T1: Auto-GEMS-TR	−2.783	8	.024	−0.928	0.449	−1.700	−0.117
Immediate Memory-ENG	-	Immediate Memory-TR	0.426	8	.681	0.142	0.226	−0.519	0.795
Months Backwards-ENG	-	Months Backwards-TR	9.071	8	< .001	3.024	1.002	1.419	4.603
Delayed Memory-ENG	-	Delayed Memory-TR	2.234	8	.056	0.745	0.489	−0.018	1.472
Naming - ENG	-	Naming-TR	−1.000	8	.347	−0.333	0.291	−0.996	0.349
Comprehension-ENG	-	Comprehension-TR	0.206	8	.842	0.069	0.534	−0.588	0.721
Orientation-ENG	-	Orientation-TR	0.000	8	1.000	0.000	0.471	−0.653	0.653
TMT-B ENG	-	TMT-B-TR	1.125	8	.293	0.375	0.379	−0.313	1.042

Note. Student's t-test.

Table 3.9. Paired sample t-test: Auto-GEMS-TR vs. Auto-GEMS-ENG

	N	Mean	SD	SE	Coefficient of variation
T1: Auto-GEMS-TR	9	87.44	8.263	2.754	0.094
T2: Auto-GEMS-ENG	9	78.89	8.100	2.700	0.103
Immediate Memory-TR	9	4.333	1.225	0.408	0.283
Immediate Memory-ENG	9	4.444	0.882	0.294	0.198
Months Backwards-TR	9	0.667	1.323	0.441	1.984
Months Backwards-ENG	9	4.667	0.500	0.167	0.107
Delayed Memory-TR	9	2.778	1.302	0.434	0.469
Delayed Memory-ENG	9	4.000	1.225	0.408	0.306
Naming-TR	9	3.556	0.527	0.176	0.148
Naming - ENG	9	3.333	0.866	0.289	0.260
Comprehension-TR	9	2.444	1.130	0.377	0.462
Comprehension-ENG	9	2.556	0.882	0.294	0.345
Orientation-TR	9	2.667	0.500	0.167	0.188
Orientation-ENG	9	2.667	0.500	0.167	0.188
TMT-B-TR	9	10.89	3.689	1.230	0.339
TMT-B ENG	9	12.56	4.333	1.444	0.345

Table 3.10. Mean scores for both conditions (Auto-GEMS-TR vs. Auto-GEMS-ENG)

The decrease in global scores across the two retest conditions is illustrated in Figure 3.3.

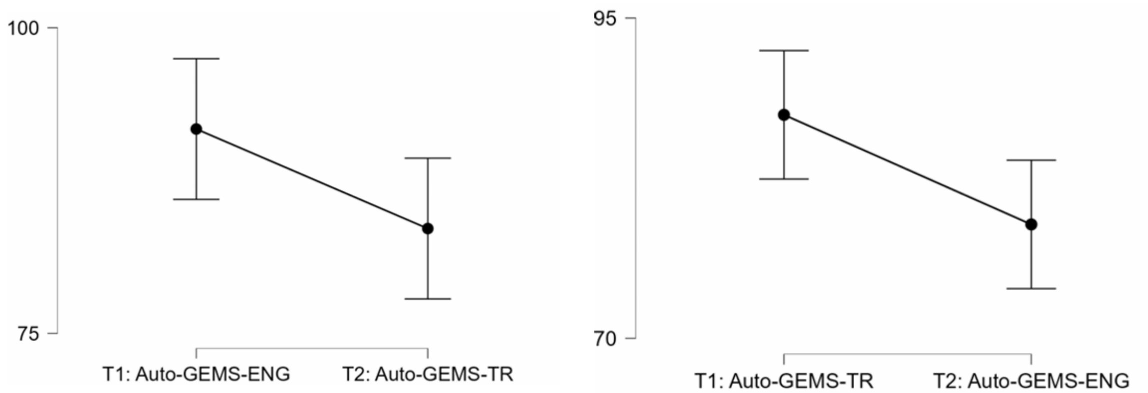

Figure 3.3. Decrease in Global Scores

All participants *(N=29)* filled out the Feedback Questionnaire presented at the end of the Turkish Auto-GEMS: 83% of respondents stated that they carried out the task to the best of their ability; 72 % stated that they carried out the task autonomously from start to finish; 28% stated that they have been interrupted or distracted during the task; 7% stated that they carried out the task quickly, without focusing on their answers; 3% stated that they carried out the task with a person who explained some of the questions to them.

CHAPTER 4: DISCUSSION & CONCLUSIONS

The primary aim of this dissertation was to collect preliminary data on the newly translated Turkish Auto-GEMS, conduct an initial analysis of the psychometric properties, and compare it to the English Auto-GEMS among bilingual individuals through a test-retest design in which a sub-sample of participants retook the alternate language version of Auto-GEMS after one month. In the present investigation, age, cognitive reserve and the level of education did not appear to influence the total score of the sample, as none of the predictors reached statistical significance and explained only a small portion of the total variation. The present study may have lacked sufficient statistical power to detect such effects due to the size of the sample, as in many cognitive tests, performance tends to decrease with age and increase with education and Cognitive Reserve (Scheffels et al., 2023; Zarantonello et al., 2020). According to the small sample of study, the internal consistency across items was poor; however, most of the items showed a significant medium-to-large positive association with the total score, showing that they still contributed meaningfully to the overall measure.

The English Auto-GEMS global score was shown to increase with the Turkish Auto-GEMS global score, but no statistically significant evidence of an association was found. Participants also seemed to perform better on the English Auto-GEMS if they had a higher language proficiency level. No practice effects were seen, but rather, a decrease in score was present for both retest conditions, no matter the language. This could point to the presence of sample-related variables, such as a decrease in participants' motivation, but also to the use of different cognitive processes for different language versions. Among the single tasks, participants significantly performed worse on the Turkish version of Months Backwards in all conditions, suggesting that this difference may be caused by linguistic ambiguities or inconsistencies rather than participant-related issues. Further investigations should be made into the linguistic properties of the test.

It is important to acknowledge that the present study has certain limitations regarding the sample. As participants were recruited via convenience sampling, the results may not be representative of the general population, especially in terms of age, education and language proficiency. Due to the small sample size, the results had limited generalizability and statistical power. Feedback given by the participants revealed additional limitations, presented with the use of online digital tests, as the results can be influenced by variables related to the test-taking process, such as distractions, low motivation, lack of supervision and guidance. The results and inferences made from the analyses should be interpreted with these limitations in mind, which should be considered in future studies. Further data collection is required to determine exactly how language affects the scores. In the future, bigger and more representative samples should be collected to focus on ways to improve the reliability and consistency across items and testing sessions. To develop a full picture of how the predictors named for the Italian version of Auto-GEMS influence the global score in the Turkish version, additional studies will be needed with more variation in the sample. Another point of inquiry could be to compare the Turkish Auto-GEMS with the Turkish MoCA to establish its validity with a bigger sample. More consideration can also be given to the other materials used in the study in relation to the Auto-GEMS, such as the MAC-Q.

Taken together, these results suggest that the Turkish Auto-GEMS, at its initial validation stage, is relatively easy to follow and understandable for participants, the majority of whom were able to complete the test autonomously. The study was successful in its attempt to conduct preliminary analyses on the psychometric properties, which also demonstrated potential limitations and areas of improvement. The methods used in this study may be applied at a larger scale to further develop the Turkish Auto-GEMS, investigate additional psychometric properties, and make cross-cultural/linguistic comparisons with other emerging language versions.

REFERENCES

Backx, R., Skirrow, C., Dente, P., Barnett, J. H., & Cormack, F. K. (2020). Comparing web-based and lab-based cognitive assessment using the Cambridge Neuropsychological Test Automated Battery: A within-subjects counterbalanced study. *Journal of Medical Internet Research*, *22*(8), e16792.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 177-196.

Bergman, I., Franke Föyen, L., Gustavsson, A., & Van den Hurk, W. (2025). Test–retest reliability, practice effects and estimates of change: A study on the Mindmore digital cognitive assessment tool. *Scandinavian Journal of Psychology*, *66*(1), 1-14.

Celik, S., Kokje, E., Meyer, P., Frölich, L., & Teichmann, B. (2022). Does bilingualism influence neuropsychological test performance in older adults? A systematic review. *Applied Neuropsychology: Adult*, *29*(4), 855-873.

Charalambous, A. P., Pye, A., Yeung, W. K., Leroi, I., Neil, M., Thodi, C., & Dawes, P. (2020). Tools for app-and web-based self-testing of cognitive impairment: systematic search and evaluation. *Journal of medical Internet research*, *22*(1), e14551.

Cicikler, B., Ozturk, H. T., & Tutar, P. (2025). What Makes Adults Digitally Literate in Türkiye: An Evaluation by Demographics. *European Journal of Education*, *60*(4), e70340.

Cipolotti, L., & Warrington, E. K. (1995). Neuropsychological assessment. *Journal of Neurology, Neurosurgery, and Psychiatry*, *58*(6), 655.

Crawford, J. R., & Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*, *20*(3), 259.

Cubillos, C., & Rienzo, A. (2023). Digital cognitive assessment tests for older adults: systematic literature review. *JMIR mental health*, *10*(1), e47487.

Cullen, B., O'Neill, B., Evans, J. J., Coen, R. F., & Lawlor, B. A. (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery & Psychiatry*, *78*(8), 790-799.

Feenstra, H. E., Murre, J. M., Vermeulen, I. E., Kieffer, J. M., & Schagen, S. B. (2018). Reliability and validity of a self-administered tool for online neuropsychological testing: The Amsterdam Cognition Scan. *Journal of Clinical and Experimental Neuropsychology*, *40*(3), 253-273.

Fernandez, A. L. (2019). Modern neuropsychological tests for a diversity of cultural contexts. *The Clinical Neuropsychologist*, *33*(2), 438-445.

Fernández, A. L., & Abe, J. (2018). Bias in cross-cultural neuropsychological testing: problems and possible solutions. *Culture and Brain*, *6*(1), 1-35.

Hansen, T. I., Haferstrom, E. C. D., Brunner, J. F., Lehn, H., & Håberg, A. K. (2015). Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors. *Journal of Clinical and Experimental Neuropsychology*, *37*(6), 581-594.

Harris, C., Tang, Y., Birnbaum, E., Cherian, C., Mendhe, D., & Chen, M. H. (2024). Digital neuropsychology beyond computerized cognitive assessment: applications of novel digital technologies. *Archives of Clinical Neuropsychology*, *39*(3), 290-304.

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-based nursing*, *18*(3), 66-67.

Huynh, D., Ye, S., Hosseini Ghomi, R., Patterson, M., & Huang, B. (2025). Reliability of remote self-administered digital cognitive assessments: preliminary validation study. *Frontiers in Digital Health*, *7*, 1571053.

Iverson, G. L., Brooks, B. L., Ashton, V. L., Johnson, L. G., & Gualtieri, C. T. (2009). Does familiarity with computers affect computerized neuropsychological test performance?. *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 594-604.

JASP (version 0.97.1; JASP Team, 2026): JASP Team (2026). *JASP* (Version 0.97.1) [Computer software]. https://jasp-stats.org/

Katzman, R., Aronson, M., Fuld, P., Kawas, C., Brown, T., Morgenstern, H., ... & Ooi, W. L. (1989). Development of dementing illnesses in an 80-year-old volunteer cohort. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *25*(4), 317-324.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American journal of health-system pharmacy*, *65*(23), 2276-2284.

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE, 10*(6), e0130834. doi.org

López, E., Steiner, A. J., Hardy, D. J., IsHak, W. W., & Anderson, W. B. (2016). Discrepancies between bilinguals' performance on the Spanish and English versions of the

WAIS Digit Span task: Cross-cultural implications. *Applied Neuropsychology: Adult*, *23*(5), 343-352.

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives of Clinical Neuropsychology*, *32*(5), 541-554.

Mondini, S., Montemurro, S., Pucci, V., Ravelli, A., Signorini, M., & Arcara, G. (2022). Global Examination of Mental State: An open tool for the brief evaluation of cognition. *Brain and Behavior*, *12*(8), e2710.

Montemurro, S., Mondini, S., Pucci, V., Durante, G., Riccardi, A., Maffezzini, S., ... & Arcara, G. (2023). Tele-Global Examination of Mental State (Tele-GEMS): an open tool for the remote neuropsychological screening. *Neurological Sciences*, *44*(10), 3499-3508.

Morrissey, S., Gillings, R., & Hornberger, M. (2024). Feasibility and reliability of online vs in-person cognitive testing in healthy older people. *PLoS One*, *19*(8), e0309006.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695-699.

Nucci, M., Mapelli, D., & Mondini, S. (2012). Cognitive Reserve Index questionnaire (CRIq): a new instrument for measuring cognitive reserve. *Aging clinical and experimental research*, *24*(3), 218-226.

Parsons, T. D. (2016). *Clinical neuropsychology and technology: What's new and how we can use it*. Springer.

Pucci, V., Contemori, G., Saccani, M. S., Arcara, G., Mondini, S., & Bonato, M. (2024). Auto-Global Examination of Mental State (Auto-GEMS): a web-based self-administered cognitive screening. *Aging Clinical and Experimental Research*, *36*(1), 216.

RStudio (version 2026.04; Posit team, 2026): R Core Team (2026). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. doi:10.32614/R.manuals https://doi.org/10.32614/R.manuals https://www.R-project.org/

Scheffels, J. F., Ballasch, I., Scheichel, N., Voracek, M., Kalbe, E., & Kessler, J. (2023). The influence of age, gender and education on neuropsychological test scores: Updated clinical norms for five widely used cognitive assessments. *Journal of Clinical Medicine*, *12*(16), 5170.

Selekler, K., Cangoz, B., & Uluc, S. (2010). Power of discrimination of Montreal cognitive assessment (MoCA) scale in Turkish patients with mild cognitive impairment and Alzheimer’s disease. *Turkish Journal of Geriatrics, 13*, 166–171.

Singh, S., & Germine, L. (2021). Technology meets tradition: a hybrid model for implementing digital tools in neuropsychology. *International Review of Psychiatry*, *33*(4), 382-393.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015-2028.

Stern, Y. (2017). An approach to studying the neural correlates of reserve. *Brain imaging and behavior*, *11*(2), 410-416.

Tsoy, E., Zygouris, S., & Possin, K. L. (2021). Current state of self-administered brief computerized cognitive assessments for detection of cognitive disorders in older adults: a systematic review. *The journal of prevention of Alzheimer's disease*, *8*(3), 267-276.

Tuerk, C., Saha, T., Bouchard, M. F., & Booij, L. (2023). Computerized cognitive test batteries for children and adolescents—a scoping review of tools for lab-and web-based settings from 2000 to 2021. *Archives of Clinical Neuropsychology*, *38*(8), 1683-1710.

Van Patten, R. (2021). Introduction to the Special Issue-Neuropsychology from a distance: Psychometric properties and clinical utility of remote neurocognitive tests. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 767-773.

Watt, J. A., Lane, N. E., Veroniki, A. A., Vyas, M. V., Williams, C., Ramkissoon, N., ... & Goodarzi, Z. (2021). Diagnostic accuracy of virtual cognitive assessment and testing: Systematic review and meta-analysis. *Journal of the American Geriatrics Society*, *69*(6), 1429-1440.

Whalley, L. J., Deary, I. J., Appleton, C. L., & Starr, J. M. (2004). Cognitive reserve and the neurobiology of cognitive aging. *Ageing research reviews*, *3*(4), 369-382.

Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: a systematic review. *Alzheimer's & Dementia*, *4*(6), 428-437.

Zarantonello, L., Schiff, S., Amodio, P., & Bisiacchi, P. (2020). The effect of age, educational level, gender and cognitive reserve on visuospatial working memory performance across adult life span. *Aging, Neuropsychology, and Cognition*, *27*(2), 302–319.