UNIVERSITY OF PADOVA

Department of General Psychology

Bachelor's Degree Course in Psychological Science

Final dissertation

Translation, Adaptation, and Validation of the Auto-Global Examination of Mental State (Auto-GEMS) for the Turkish Population

***Supervisor* Professor**
Professor Dr. Mario Bonato

Co-supervisor
Vincenzo Livoti

Candidate: *Derin Akcan*

Student ID number: *2081282*

Academic Year 2024-2025

ABSTRACT

This study presents the translation, cultural adaptation, and preliminary validation of the Auto-Global Examination of Mental State (Auto-GEMS) for the Turkish population. Auto-GEMS is a self-administered, web-based cognitive screening tool composed of 11 subtests targeting various cognitive domains, as well as an adapted Cognitive Reserve Index Questionnaire (CRIq) to assess cognitive reserve. The translation process included translation and back-translation, cross-linguistic analyses based on word frequency, imageability, and semantic similarity to ensure comparability between the Italian and Turkish versions. Following adaptation, a pilot validation study was conducted to assess the clarity, usability, and preliminary validity of the Turkish version.

INTRODUCTION

The growing interest in cross-cultural neuropsychological research is reflected in the increasing number of studies focused on translating cognitive screening instruments. Such translations can take various forms: an instrument may be directly translated from one language to another, or, in cases of significant cultural differences, the process may involve adapting the tool or even developing an entirely new version that assesses the same construct (Vijver et al., 1996). This study provides a theoretical background by first introducing the concept of digital neuropsychological testing and discussing current translation and adaptation practices, which often remain insufficient. It then presents Auto-GEMS, a digital cognitive screening tool that was translated and adapted into Turkish. The materials used, including Auto-GEMS and accompanying instruments, are described in detail, followed by an explanation of the translation and cultural adaptation process, which was conducted based on the International Test Commission (ITC) guidelines. Finally, a pilot validation study was conducted to collect qualitative feedback and perform preliminary statistical analyses.

CHAPTER 1: THEORETICAL BACKGROUND

1.1 DIGITAL NEUROPSYCHOLOGICAL TESTING

Digital neuropsychology is defined as: “the neuropsychological assessment of cognition and behavior using digital tools.” (Germine et al., 2019) and it can take a variety of forms, ranging from proctored examinations to fully online, self-administered screenings. However, before exploring the digital domain of cognition, it is important to distinguish between neuropsychological testing and neuropsychological assessment. According to the joint statement of the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) (Bauer et. al, 2012), neuropsychological testing refers to the usage of cognitive tests to obtain behavioral samples in domains including memory, concentration, or executive function; while neuropsychological assessment involves the integration of patient history, symptoms, behavioral observations, physical findings, and other relevant information about the patient's situation to infer the underlying cause of the performance pattern. Under this definition, self-administered online cognitive screenings such as Auto-GEMS are best classified as neuropsychological tests rather than assessments. These tests can be administered by technical staff or through self-assessment. However, the results must still be interpreted by trained professionals. AACN and NAN stress that this task should remain in the hands of experts in psychometrics and clinical neuropsychology.

Having clarified the distinction between neuropsychological testing and assessment, the discussion can now turn to the distinctive features of digital cognitive testing. Rather than proposing it as a replacement of traditional cognitive assessments, it is more appropriate to consider the contexts in which online cognitive tests are most suitable. One of the most important qualities of online testing is the possibility to assess and monitor cognitive function remotely. Remote neurocognitive testing exists on a continuum. At one end is computerized administration via videoconferencing. In the middle are structured but unsupervised tests, such as Auto-GEMS. At the other end is mobile cognitive testing, which uses smartphones or other portable devices for repeated and ambulatory assessments (Ryan Van Patten, 2021).

Figure 1 *The continuum of remote cognitive testing*

Note. From Ryan Van Patten (2021)

With the aging population, there is a growing need for valid, easily accessible tools to assess cognitive function and monitor cognitive decline. Remote assessment offers several advantages: it reduces costs by minimizing the need for staff time, travel, and equipment, and it enhances accessibility by reaching larger and more diverse populations (Cambridge Cognition, 2020). This is particularly relevant for Turkey. In a study that conducted a cluster analysis of Turkish regions based on health indicators, Köse (2022) found that Southeastern, Northeastern, and Middle Eastern Anatolia consistently clustered together as regions with lower performance in health service demand, production, and capacity. Therefore, implementing remote neuropsychological testing in low-SES and rural areas is especially important. These regions have limited access to clinical facilities, and remote testing can significantly improve both accessibility and representation in cognitive health research and care.

Despite the advantages of remote neuropsychological testing, there are important drawbacks to consider. One major limitation is the lack of a standardized clinical environment, which prevents clinicians from fully controlling the testing conditions (Cambridge Cognition, 2020). Remote assessments are subject to variability in participants' hardware, software, and internet connectivity, all of which can affect test performance and validity of the results. These challenges are especially pronounced in lower SES and rural regions of Turkey, where access to computers and stable internet connection is limited.

According to Yıldız (2009), the “digital divide” in Turkey is shaped by socioeconomic and geographic disparities, with rural areas demonstrating significantly lower access to Information and Communication Technology resources compared to urban areas. High internet costs, lack of infrastructure, and limited digital literacy further restrict access, particularly for women, the elderly, and those with low educational attainment (Yıldız 2009). Given these factors, it is important to interpret test scores by taking participants’ digital literacy into account.

1.2 TRANSLATING COGNITIVE SCREENING TESTS

In the context of cross-cultural research, a key concern is ensuring that cognitive screening tools maintain their validity and reliability across different populations. This necessity highlights the importance of standardized guidelines for translating and adapting such instruments. Vijver et al. (1996) identifies three major biases in translation research: construct bias, method bias, and item bias. Construct bias arises when a measured construct has notable differences across cultures –such as intelligence, which is often viewed holistically in non-Western cultures and more logically and analytically in the West (American Psychological Association, 2003). Method bias refers to environmental, procedural, and interpersonal factors during testing that may distort validity if not properly controlled for. Item bias classifies errors in test items such as inaccurate translations, unclear wording, and culturally inappropriate content, that may lead to disadvantage in certain cultural groups.

A later paper published by the International Test Commission (ITC, 2017) presents a detailed set of guidelines for translation and adaptation procedures that help to minimize these biases. It offers several suggestions under each guideline aiming to solve a variety of complications that may occur throughout the process. In the context of our study, ITC guidelines emphasize the procedures of forward and back-translation, inclusion of experts within the domains of language, culture, and psychology. Experts are preferred to be bilingual in order to ensure fluency and an indigenous appearance of the translated version. Additionally, the ITC strongly recommends the collection of pilot data to run small scale validity and reliability tests and conduct item analysis, which makes it possible to make the necessary changes before finalizing the translated version. Beyond these development procedures, the ITC presents a section dedicated to empirical validation. It recommends selecting representative and sufficiently large samples, and conducting both exploratory and confirmatory factor analyses to examine the factor structure across the original and translated versions. These analyses help to determine whether the underlying construct is measured consistently across languages, thus ensuring that the translation is highly comparable to its source version (Fenn et al., 2020). Finally, the guidelines calls for a transparent documentation process of all stages of the adaptation and validation process, including rationale for translation decisions and outcomes of statistical analyses.

Despite the existing literature presenting detailed instructions for translation and adaptation, studies reveal that even the most widely used cognitive assessment tests' translated versions often fail to meet the aforementioned criteria. A systematic review examining linguistic and cultural adaptations of the MoCA (Montreal Cognitive Assessment) reported that only 2 out of 86 versions followed all recommended steps such as translation and back-translation, professional translators, involvement with the original author, and pilot testing (Cova et al., 2022). Furthermore, many of the translations lacked detailed documentation of their process. Another study reviewing the translated and culturally adapted versions of the MMSE (Mini-Mental State Examination) in 15 languages showed that the validity of the translations varied widely, with some of them being rigorously validated while others lacking reliable data (Steis & Schrauf, 2009).

1.3 AUTO-GEMS

Auto-Global Examination of Mental State (Auto-GEMS), developed by Pucci et al (2024), is a self-administered, web-based cognitive screening test. It includes two parallel versions: Auto-GEMS A and Auto-GEMS B. The two versions share the same structure and differ only in item content. Each version comprises 11 tasks, with each task assessing a different aspect of cognitive function. Auto-GEMS also evaluates cognitive reserve using a short, adapted questionnaire based on the Cognitive Reserve Index Questionnaire (CRIq), developed by Nucci et al. (2012). In addition to these measures, basic demographic and user experience data is collected.

1.3.1 Items

Each item was scored individually and one point awarded for each correct response. The 11 subtests are the following:

1) *Orientation:* Assesses temporal and spatial orientation through questions about the current season, year, and the geographic location of a place relative to another.
2) *Immediate Memory:* Evaluates short-term verbal memory by presenting six words both visually and auditorily, followed by a free recall task.
3) *Months Backwards:* Measures working memory. Participants are asked to recite months in reverse order and skip one each time. (e.g., October, August, June).

4) *Puzzle Task:* Tests visuoconstructional skills by instructing the participant to reconstruct a fragmented image using drag-and-drop.
5) *Clock task:* Assesses spatial reasoning by asking participants to indicate the positions of clock hands in imagined configurations.
6) *Delayed Memory:* Examines long-term verbal memory through the delayed recall of the previously presented six words.
7) *Naming:* Tests language and lexical retrieval by asking participants to name non-living objects presented as images.
8) *Comprehension:* Measures the understanding of simple commands (verbal and written) by asking the participant to press two letters after pressing another.
9) *Trail Making Test A (TMT-A):* Assesses attention and processing speed. Participants click on numbered circles in ascending order.
10) *Trail Making Test B (TMT-B):* Tests divided attention and cognitive flexibility through alternating sequences of numbers and letters.
11) *Metaphor Comprehension:* Evaluates pragmatic language skills by presenting a sentence with a figurative meaning and asking for its correct interpretation among multiple options.

1.3.2 Cognitive Reserve Index (CRI)

Stern (2009) defines cognitive reserve (CR) as "individual differences in the cognitive processes or neural networks underlying task performance that allow some people to cope better than others with brain damage." The strongest predictors of cognitive reserve have been shown to be education and occupational attainment (Stern, 2002). Cognitive reserve has a strong positive correlation with cognitive performance (Opdebeeck et al, 2015). Therefore, Auto-GEMS accounts for cognitive reserve by using the shortened version of the Cognitive Reserve Index Questionnaire developed by Mondini et al. (2023). This questionnaire collects information about participant's education level, working activity, and leisure time. For education level, it questions the highest degree earned and the duration of any additional courses taken. Regarding working activity, it includes the type and number of years of both primary and secondary jobs, if any. For leisure time, it asks whether the participant engages in volunteering, cultural activities (such as studying, reading books, or visiting museums), socialization, and sports. For each of these, it records how many times per week the activity is done. These three components are averaged to compute a total cognitive reserve score, which quantifies cognitive reserve for each participant. This score can be applied to multiple regression models as a predictor variable to explain performance in Auto-GEMS. As Pucci et al. (2024) demonstrated in the Italian version of Auto-GEMS, cognitive reserve had a significant positive effect ($\beta = 0.05$, $p = 0.002$) on the total Auto-GEMS score, showing that higher cognitive reserve is associated with better performance.

1.3.3 Cut - offs

In the Italian version of Auto-GEMS, the cut-off scores were calculated by using the Crawford and Garthwaite regression-based approach (Crawford and Garthwaite, 2006). This method predicts expected performance from an individual's age, education, sex, and cognitive reserve score, allowing for a more individualized interpretation of whether their score is within normal limits or not.

CHAPTER 2: METHODOLOGY

2.1 MATERIALS

The following materials were used in the study: a brief medical history form, the MAC-Q, the CRI-Q, Auto-GEMS-A, and a post-assessment questionnaire, all translated to Turkish (see Appendix II).

Brief Medical History Form: The medical history form collected information about participants' current or past health issues that may affect cognitive function. It included items on disabilities, neurological events, medication use, and other relevant health concerns.

MAC-Q (Memory Complaint Questionnaire): MAC-Q is a short self-report memory questionnaire developed by Crook et al. (1992) to assess age-associated memory complaints.

CRI-Q (Cognitive Reserve Index Questionnaire): As described in Section 1.3.2, CRI-Q was administered to measure cognitive reserve by evaluating factors such as education, working activity, and leisure time.

Auto-GEMS-A: The Turkish translation and adaptation of Auto-GEMS version A was used.

Post-Assessment Questionnaire: This questionnaire was done at the end of the session. Participants were asked about their testing experience, including whether they experienced distractions, whether they completed the test independently, or if others assisted them in accessing or completing the online task.

2.2 TRANSLATION AND ADAPTATION

Medical history form, MAC-Q, CRI-Q and post-assessment questionnaire were all directly translated from English to Turkish. The vocal instructions were converted into Turkish using artificial intelligence voice synthesis from the website https://ttsmp3.com/. For Auto-GEMS itself we followed a rigorous procedure grounded on forward-and-back translation and psycholinguistic variables such as word frequency and imageability.

2.2.1 Translating and adapting the items and instructions

A forward-and-back translation method was adopted during the translation of Auto-GEMS items and instructions. Two groups were formed: the forward and back-translation group. The forward translation group translated the introduction, comprehension, working memory, immediate memory, puzzle, spatial, TMT (A and B) and metaphor tasks. These translations were then sent to the back-translation group. The back-translation group worked to translate the document back into English. In order to eliminate bias, the back-translation group did not communicate with the translation group nor did they read the official English text. We ensured that our translation of the Turkish text into English was as direct as possible, making only minor punctuation adjustments to preserve the context. Upon the completion of the back-translation, a judge with no prior exposure to the translations was assigned to review and compare the original English text with the back-translated English version. In this step, we focused on ensuring that the meaning of the content stayed the same as the original English version within the back-translation. Additionally, clarity of content and correct use of terminology were also controlled by the judge. The final translations were refined based on the judge’s feedback.

In the orientation task, one of the three questions involves identifying the geographical location of one place relative to another. In Version B of the test, the item asks: "Relative to Asia, where is Europe located?" Since this question refers to continents and is likely to be universally understood, it was translated directly without modification. However, in Version A, the original item asks: "Relative to Venice, where is Rome located?" Given that this question references specific Italian cities, we adapted it to include two well-known Turkish cities instead.

2.2.2 Translating and adapting the word recall task

The procedure of the adaptation of the word recall task is based on three concepts: word frequency, imageability, and semantic similarity. In psycholinguistics, *frequency* refers to the number of occurrences of a particular linguistic item in a corpus. *Imageability* is a psycholinguistic variable that indicates how well a word gives rise to a mental image or sensory experience (Rofes et al., 2018, p. 1188). *Semantic similarity* refers to the comparison of meaning between words or concepts (Hahn & Heit, 2015). Several factors explain why we have picked these variables as the basis of our adaptation. Firstly, word frequency is one of the most important determinants of word recall performance (Hulme et al., 1997; Poirier & Saint-Aubin, 1996). Therefore, we controlled this variable by ensuring that the word frequencies in the Turkish and Italian versions of the test are concordant. Secondly, imageability has been found to influence word recall task outcomes (Tse & Altarriba, 2007). Lastly, studies show potential confounding effects due to semantic variations between languages (Lupyan et al., 2018). Thus, in addition to keeping the meanings of the Turkish and Italian versions as similar as possible, we took into account frequency values and imageability to ensure that the tasks are highly comparable, both across languages and between version A and B.

First, we examined the dataset used in the word selection process for the Italian version of Auto-GEMS; SUBTLEX-IT, a frequency database for Italian words based on movie subtitles (Amenta et al., 2024). Several other SUBTLEX databases exist for different languages. This particular database provides subtitle frequency estimates for Italian, based on approximately 128 million words. For the Turkish version, we selected the OpenSubtitles Corpus for Turkish (Lison & Tiedemann, 2016), which we found on the Open Parallel Corpora (OPUS) website (Tiedemann, 2012). This dataset contains approximately 151 million words. For frequency comparison, we used the Frequency Profiling Method, proposed by Rayson and Garside (2002). This method involves calculating the log-likelihood statistic for each word in the two frequency lists, achieved by constructing a contingency table as in Table 1.

Table 1 *Contingency table for word frequencies*

	SUBTLEX–IT	**OPUS**	**TOTAL**
Freq. of word	a	b	$a + b$
Freq. of other words	$c - a$	$d - b$	$c + d - a - b$
TOTAL	c	d	$c + d$

The value “c” corresponds to the number of words in SUBTLEX-IT, and “d” corresponds to the number of words in OPUS. The values of “a” and “b” are called the observed values (O). We calculated the expected values (E) according to the following formula:

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

In our case, $N_1 = c$ and $N_2 = d$ So for a word, $E_1 = c \times (a + b) \div (c + d)$ and $E_2 = d \times (a + b) \div (c + d)$. We then calculated the log-likelihood value according to this formula:

$$- 2 \ln \lambda = 2 \sum_i O_i \ln\left(\frac{O_i}{E_i}\right)$$

This equates to calculating LL as follows:

$$LL = 2 \times \left(\left((a \times log\left(\frac{a}{E_1}\right)\right) + \left(b \times log\left(\frac{b}{E_2}\right)\right)\right)$$

The log-likelihood value represents the semantic tag’s frequency deviation from the normative corpus, which in our case is SUBTLEX-IT. The higher the LL value, the greater the deviation. To interpret the LL values, we did not conduct a hypothesis test, as the distribution of words within the corpora is not random. Instead, we qualitatively evaluated and matched the words based on their relevance and similarity.

Parallel to this procedure, in order to minimize the influence of word imageability on word recall task accuracy, we selected words with high imageability from two databases compiled by Selvi-Balo et al. (2020) and Acar et al. (2016). Both studies measured imageability by having participants rate each word on a 7-point Likert scale, with 1 representing the lowest and 7 the highest imageability. According to Selvi-Balo et al. (2020) values at or below 5.955 were classified as "low" for imageability based on the 25th percentile, while values at or above 6.713 and were classified "high" based on the 75th percentile. According to our statistical analysis of the dataset by Acar et al. (2016), we found the 25th and 75th percentiles to be 3.86 and 6.16, respectively. Based on these numbers, we selected words from both databases with a median imageability score greater than the 75th percentiles.

By applying the aforementioned frequency profiling method, we calculated the frequency values for each potential Turkish word by coding the formulas in R studio, and selected those that minimized the log-likelihood value when compared to their Italian counterparts (see Appendix I). The resulting frequency values were 14,000, 11,800, 1,100, 1,000, 930, and 630. Then, using words derived from the two imageability databases, we compiled a sub-list of words whose frequency values in the OPUS corpus aligned with the calculated frequency values.

Semantic similarity refers to the comparison of meaning between words or concepts (Hahn & Heit, 2015). To ensure semantic similarity between Turkish and Italian words, we selected words that closely matched those in the Italian version (see Appendix). We aimed to keep the semantic similarity as close as possible between the versions A and B, and as different as possible within each version as the original Italian Auto-GEMS study. This process aimed to preserve the conceptual structure and lexical characteristics of the original tasks, ensuring that the cognitive demands remain comparable across languages.

The final version of selected words for the Auto-GEMS Turkish Version - A & B are illustrated in Table 2 and 3.

***Table 2** Selected words for version A*

TUR–A	FREQ (tur)	ENG
Güneş	13569	Sun
Kulak	10491	Ear
Yastık	1029	Pillow
Kibrit	770	Match
Çatal	740	Fork
Yelek	687	Vest

***Table 3** Selected words for version B*

TUR–B	FREQ (tur)	ENG
Yıldız	10233	Star
Parmak	9631	Finger
Gözlük	1101	Glasses
Cüzdan	691	Wallet
Kaşık	960	Spoon
Kazak	777	Sweater

2.3 PARTICIPANTS AND PROCEDURE

Participants were recruited using convenience sampling. Inclusion criteria included being over the age of 18, while participants with a history of neurological or psychiatric conditions were excluded. A recruitment flyer was shared via Instagram and various WhatsApp groups. The flyer included a Google Drive link that provided access to the study instructions and the test itself. In the instructions, participants were given a brief explanation of Auto-GEMS and informed about the additional questionnaires they would be asked to complete. The purpose of the study was also stated clearly: participants were told that this was a preliminary validation study and that their feedback would be highly valuable.

Participants were first instructed to create a personalized code combining their place of birth and age (e.g., ISTANBUL29) to ensure anonymity throughout the study. They then completed a medical history questionnaire along with the Memory Assessment Clinics Questionnaire (MAC-Q), both hosted on Qualtrics. Upon submission, they were automatically redirected to the Auto-GEMS cognitive test, administered via the JATOS platform. JATOS (Just Another Tool for Online Studies), developed by Lange et al. (2015) is an open-source platform designed to run and manage online behavioral experiments. It can be run locally or on a server. Auto-GEMS was imported on the University Of Padova JATOS server, which enabled it to generate a single-use link and collect individual data remotely.

An informed consent form was presented and those who agreed to proceed were asked to re-enter their personalized ID code and provide basic demographic information. Following this, the Cognitive Reserve Index questionnaire was presented. After completing all preliminary measures, participants proceeded to the test. After completing Auto-GEMS, participants answered a brief questionnaire regarding the testing conditions. They were asked whether they completed the test independently, if they received any assistance, and whether there were any distractions during the session. A feedback box was also provided at the end of the session, allowing participants to comment freely on their experience. Data from the Auto-GEMS test, including the CRI scores, questionnaire responses, and open-ended feedback, were collected via the JATOS platform.

CHAPTER 3: RESULTS

3.1 QUALITATIVE FEEDBACK

A total of 51 participant feedback responses were collected. The majority of participants reported that the instructions and the overall test were clear. Seventeen responses contained negative or constructive feedback. Purely negative responses (e.g., "the test wasn't clear") were not analyzed in depth; instead, emphasis was placed on constructive suggestions that could inform future improvements.

One of the most frequently mentioned concerns was related to the lack of clarity regarding score interpretation. Several participants expressed confusion or anxiety about what their scores meant and whether their performance was considered “good” or “bad.” Some reported feeling insecure or stressed as a result. Additionally, issues related to anonymity were raised by two participants. One suggested more strongly emphasizing that responses are anonymous, while another criticized the use of place of birth and age to create participant codes, especially given that sensitive medical information was collected. Feedback also pointed out the importance of pre-test instructions regarding mental and physical readiness. One participant, for instance, recommended explicitly stating that the test should be taken while alert, where one can direct their focus entirely on the task.

Regarding specific test items, two participants found the metaphor task to be poorly worded, leading to confusion about whether the question referred to the full sentence or a specific word. Two others noted that the orientation task was open to interpretation, making it unclear how answers were being judged. Finally, some participants mentioned an issue in the

TMT task: after selecting the letter "C", they instinctively looked for the letter "Ç", which follows "C" in the Turkish alphabet. This caused momentary hesitation and may have impacted response times.

3.2 QUANTITATIVE ANALYSIS

A total of 94 Turkish individuals fully completed the test. All participants were native Turkish speakers. The mean age of the sample was 44.9 years (SD = 17.6), with ages ranging from 18 to 78. The mean years of education was 16.2 (SD = 2.3), within a range of 13 to 19 years. The mean cognitive reserve score, as measured by the Turkish version of Auto-GEMS-A, was 132.3 (SD = 18.6), ranging from 99 to 186. The mean total Auto-GEMS-A performance score across participants was 78.9 out of 100 (SD = 10.9), with scores ranging from 55 to 100. The distribution of scores was slightly left-skewed (skewness = –0.23), indicating a tendency toward higher performance scores. Further descriptive statistics are presented in Table 4.

Table 4 *Descriptive statistics of the sample*

	mean	sd	median	min	max	kurt	skew
Age	44.891	17.604	50	18	78	-1.481	-0.22
Education	16.274	2.257	16	13	19	-1.316	-0.282
CR	132.274	18.605	130	99	186	-0.127	0.527
Orientation	1.935	0.509	2	1	3	0.721	-0.111
Immediate Memory	4.065	1.488	4	1	6	-0.997	-0.15
Working Memory	3.022	1.95	3	0	5	-1.936	-0.039
Puzzle	2.989	0.104	3	2	3	85.065	-9.281
Spatial	3.793	0.584	4	1	4	10.626	-3.21
Delayed Memory	3.565	1.66	4	0	6	-0.633	-0.242
Naming	3.75	0.527	4	1	4	7.242	-2.426
Comprehension	2.533	0.931	3	0	3	1.78	-1.79
TMT-A	13.728	1.846	14	0	14	42.616	-6.598
TMT-B	13.315	2.358	14	0	14	12.833	-3.555
Metaphor	0.663	0.475	1	0	1	-1.556	-0.679
Auto-GEMS score	78.924	10.946	79	55	100	-0.787	-0.225

An exploratory factor analysis (EFA) was done to see the structure of the cognitive test scores and check if it was similar to the original Auto-GEMS. The sample size was slightly smaller than the recommended number of 10 participants per subtest (110 participants; Fenn et al., 2020). Maximum Likelihood extraction was used without rotation. Two factors were kept based on eigenvalues greater than 1. The first factor had high loadings for Immediate Memory (.892), Delayed Memory (.869), and Working Memory (.742), showing it is related to memory. The second factor had strong loadings for Trail Making Test Part B (TMT-B; .849) and Part A (TMT-A; .524), which relate to executive function and processing speed. The eigenvalues for the first and second factors were 2.48 and 1.78, explaining 17.9% and 11.1% of the total variance. The model fit was good, with a chi-square test showing no significant difference between the model and the data, $\chi^2(34) = 31.29$, $p = .601$. Details are presented in Table 5 and 6.

Table 5 *Factor loadings*

	Factor 1	Factor 2	Uniqueness
immediate_memory_correct	0.892		0.200
delayed_memory_correct	0.869		0.236
working_memory_correct	0.742		0.449
TMTB_correct		0.849	0.254
TMTA_correct		0.524	0.725
comprehension_correct			0.891
naming_correct			0.970
orientation_correct			0.952
spatial_correct			0.963
puzzle_correct			0.997
metaphor_correct			0.968

Note. No rotation method applied.

Table 6 *Factor* characteristics

	Eigenvalues	SumSq. Loadings	Proportion var.	Cumulative
Factor 1	2.476	2.172	0.197	0.197
Factor 2	1.775	1.224	0.111	0.309

CHAPTER 4: DISCUSSION

During the review of test data, two subtests emerged as problematic: the metaphor comprehension and the orientation tasks. Specifically, the metaphor comprehension item, which presented the sentence “I visited the city library today, very rich archive!”, yielded a high error rate. Out of 94 participants, 31 selected the incorrect option, most frequently choosing “Archive is in the city library.” This suggests that the instructions failed to reflect the aim of the task. Two possible modifications could address this issue. First, the task prompt could be reformulated to directly target the metaphorical component by asking participants what the word “rich” implies in this context. Second, an alternative metaphor may be selected altogether. Similarly, the orientation task raised interpretive challenges. One item asked: “Relative to Ankara, where is Diyarbakır located?”. Diyarbakır is geographically located in the southeast of Turkey and Ankara is situated centrally. Participant responses were clustered around “south” and “east”. A simple solution would be to select city pairs that are more unambiguously oriented along the same latitude or longitude.

A further technical issue was identified in the word recall task. Five of the six target words included Turkish-specific characters such as “ş,”“ç,” or “ü.” As the current scoring system does not recognize their latin equivalents (e.g., “gunes” instead of “güneş”), responses typed without Turkish keyboard support were marked incorrect. This issue is unlikely to reflect participants' memory performance and more likely due to technological limitations. To mitigate this, the scoring algorithm can be adjusted to accept latin alternatives as correct responses. A culturally specific issue also emerged in the Trail Making Test (TMT). In the Turkish alphabet, the letter “Ç” follows “C,” and several participants expected to find “Ç” after selecting “C” during the TMT task. The absence of this character caused hesitation, potentially influencing their response times. Rather than modifying the letter sequence, a simple instruction note at the beginning of the task could clarify that the sequence will follow the English alphabet (i.e., A–B–C–D, etc.) and exclude non-Latin Turkish characters such as "Ç".

The memory tasks (Immediate Memory, Delayed Memory, and Working Memory) loaded strongly onto the first factor, while the Trail Making Test components (TMT-A and TMT-B) formed the second factor. This structure is consistent with the Italian validation study of Auto-GEMS, although differences in factor loading patterns were observed. In the Italian study, TMT-A and TMT-B exhibited the strongest loadings overall, and the second factor comprised primarily immediate and delayed memory. These differences may stem from disparities in sample size and demographic characteristics. Importantly, the exploratory factor analysis produced promising results.Whereas the Italian validation relied on a well-controlled sample of over 1,000 individuals, the present study's sample size was smaller and more heterogeneous, limiting generalizability and statistical power.

Potential limitations were regarding the overall sample. It is acknowledged that it was not representative since participants were recruited via convenience sampling. As such, the sample may not reflect the broader Turkish population in terms of age, education level, or geographic diversity. Despite these limitations, the study successfully achieved its primary goal: translating, adapting, and piloting Auto-GEMS in Turkish. The results highlight both strengths of the adapted version and areas in need of improvement, which will benefit later studies that investigate reliability and validation of Auto-GEMS for the Turkish population.

CONCLUSION

The present study has translated and culturally adapted the original version of Auto-GEMS A and B. It then conducted a pilot testing of the Turkish version of Auto-GEMS A and conducted qualitative feedback and a preliminary qualitative analysis. The analyzed constructive feedbacks and the quantitative analysis indicated a need for revising the metaphor comprehension, orientation, word recall, and TMT tasks. Factor analysis revealed a similar pattern to that of the Italian version. Limitations were regarding the size and the controlling of sampling. Future studies can better control the sampling process together with implementing more in depth statistical analyses to enhance validity and reliability. Since a detailed analysis of additional materials (MAC-Q; CRI-Q, Medical History Form, Post-Assessment Questionnaire) used in this study requires a bigger scope, similar future studies can also take this into consideration while conducting studies in the same field.

Bibliography

Acar, E. A., Zeyrek, D., Kurfalı, M., & Bozşahin, C. (2016). A Turkish database for psycholinguistic studies based on frequency, age of acquisition, and imageability. Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016), 3600–3606. European Language Resources Association (ELRA).

Amenta, S., Mandera, P., Keuleers, E., Brysbaert, M., & Crepaldi, D. (2024). SUBTLEX-IT. OSF. https://osf.io/zg7sc

American Psychological Association. (2003, February 1). Intelligence across cultures. Monitor on Psychology, 34(2). https://www.apa.org/monitor/feb03/intelligence

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. https://doi.org/10.1080/13854046.2012.663001

Cambridge Cognition. (2020, October 1). Remote cognitive testing and lessons learned during COVID-19 [Video]. YouTube. https://www.youtube.com/watch?v=3WnfQGDRjXY

Cova, I., Nicotra, A., Maestri, G., et al. (2022). Translations and cultural adaptations of the Montreal Cognitive Assessment: A systematic and qualitative review. Neurological Sciences, 43, 113–124. https://doi.org/10.1007/s10072-021-05716-y

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.

Fenn, J., Tan, C.-S., & George, S. (2020). Development, validation and translation of psychological tests. BJPsych Advances, 26(5), 306–315. https://doi.org/10.1192/bja.2020.33

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. The Clinical Neuropsychologist, 33(2), 271–286. https://doi.org/10.1080/13854046.2018.1535662

Hahn, U., & Heit, E. (2015). Semantic similarity, cognitive psychology of. In International Encyclopedia of the Social & Behavioural Sciences (2nd ed., pp. 579–584). https://doi.org/10.1016/B978-0-08-097086-8.53026-8

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5), 1217–1232. https://doi.org/10.1037/0278-7393.23.5.1217

JASP Team (2024). JASP (Version 0.19.3)[Computer software].

Köse, A. (2022). Türkiye'de sağlık göstergelerine göre istatistik bölgelerin kümeleme analizi ile sınıflandırılması [Classification of statistical regions by clustering analysis according to health indicators in Turkey]. Alanya Akademik Bakış, 6(2), 2167–2189. https://doi.org/10.29023/alanyaakademik.1021019

Lange K, Kühn S, Filevich E (2015) Correction: "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. PLoS ONE 10(7): e0134073. https://doi.org/10.1371/journal.pone.0134073

Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA).

Mondini, S., Pucci, V., Pastore, M., et al. (2023). s-CRIq: The online short version of the Cognitive Reserve Index Questionnaire. Aging Clinical and Experimental Research, 35, 2903–2910. https://doi.org/10.1007/s40520-023-02561-1

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, 218–226. https://doi.org/10.3275/7800

Opdebeeck, C., Martyr, A., & Clare, L. (2015). Cognitive reserve and cognitive function in healthy older people: A meta-analysis. Aging, Neuropsychology, and Cognition, 23(1), 40–60. https://doi.org/10.1080/13825585.2015.1041450

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity, and item position. Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 50(4), 408–412. https://doi.org/10.1037/1196-1961.50.4.408

Pucci, V., Contemori, G., Saccani, M. S., Arcara, G., Mondini, S., & Bonato, M. (2024, November 20). Auto-Global Examination of Mental State - Auto-GEMS. OSF. https://doi.org/10.17605/OSF.IO/FQ8G7

Rayson, P., & Garside, R. (2002). Comparing corpora using frequency profiling. In Proceedings of the Workshop on Comparing Corpora (pp. 1–6). Association for Computational Linguistics. https://doi.org/10.3115/1117729.1117730

Reid, M., Parkinson, L., Gibson, R., Schofield, P., D'Este, C., Attia, J., Tavener, M., & Byles, J. (2012). Memory Complaint Questionnaire performed poorly as screening tool: Validation against psychometric tests and affective measures. Journal of Clinical Epidemiology, 65(2), 199–205. https://doi.org/10.1016/j.jclinepi.2011.06.006

Rofes, A., Zakariás, L., Ceder, K., Lind, M., Johansson, M. B., de Aguiar, V., ... Howard, D. (2018). Imageability ratings across languages. Behavior Research Methods, 50(3), 1187–1197. https://doi.org/10.3758/s13428-017-0936-0

Selvi-Balo, S., Maviş, İ., & Tunçer, A. M. (2020). 586 Türkçe sözcüğün imgelenebilirlik, tanıdıklık ve öznel edinim yaşı norm değerleri. Dil, Konuşma ve Yutma Araştırmaları Dergisi, 3(3), 301–334.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. Journal of the International Neuropsychological Society, 8(3), 448–460. https://doi.org/10.1017/S1355617702813248

Stern, Y. (2009). Cognitive reserve. Neuropsychologia, 47(10), 2015–2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

Thienel, R., Borne, L., Faucher, C., Robinson, G. A., Fripp, J., Giorgio, J., Ceslis, A., McAloney, K., Adsett, J., Galligan, D., Martin, N. G., Breakspear, M., & Lupton, M. K. (2023). Can an online battery match in-person cognitive testing in predicting age-related cortical changes? medRxiv. https://doi.org/10.1101/2023.04.24.23289014

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012) (pp. 2214–2218). https://aclanthology.org/L12-1246/

Tse, C. S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. Memory, 15(6), 675–686. https://doi.org/10.1080/09658210701467186

Van de Vijver, F., & Hambleton, R. (1996). Translating tests. European Psychologist, 1, 89–99. https://doi.org/10.1027/1016-9040.1.2.89

Yildiz, M. (2009). Digital divide in Turkey: A general assessment. In E. Ferro, Y. K. Dwivedi, J. R. Gil-Garcia, & M. D. Williams (Eds.), Handbook of research on overcoming digital divides: Constructing an equitable and competitive information society (pp. 75–89). IGI Global. https://doi.org/10.4018/978-1-60566-699-0.ch005

Appendices

Appendix I

Italian Word Frequencies in Auto-GEMS A & B Word Recall Task (based on SUBTLEX-IT)

ITA – A	FREQ.	ITA – B	FREQ.
Sole	9987	Torta	5258
Chiodo	531	Dito	3869
Bocca	11851	Lago	3003
Arco	963	Luna	6539
Faro	788	Busta	1844
Prato	944	Filo	2762

Appendix II

Section A: Medical History form in English and Turkish

PARTICIPANT MEDICAL HISTORY FORM

Participant n. ________ **Test date**: ___________________

GENERAL INFORMATION ABOUT THE PARTICIPANT

Surname and Name: ____

Date of Birth: _______________________Age: ______________________

Education (years of schooling): ________________________________

Male □ Female □ Right-handed □ Left-handed □ Ambidextrous □

Occupation: ____

(if retired, please indicate the job before retirement)

HEALTH INFORMATION

Do you have visual and/or hearing impairments? YES □ NO □

If yes, please specify and state whether they are corrected:

____

Do you have any health problems? YES □ NO □

If yes, please specify the type:

____

Have you ever had a stroke? YES □ NO □

If yes, please specify the details:

____

Have you ever had a head injury? YES □ NO □

If yes, please specify the details:

____

Have you ever had epileptic seizures or convulsions? YES □ NO □

If yes, please specify the details:

____

1

Have you ever needed a neurological consultation? YES □ NO □
If yes, please specify the details:

____

Have you ever needed a psychological or psychiatric consultation? YES □ NO □
If yes, please specify the details:

____

Has any of your family members ever needed a psychological, psychiatric, or neurological consultation? YES □ NO □
If yes, please specify the details:

____

Have you ever used or are currently using drugs and/or abusing alcohol? YES □ NO □
If yes, please specify the details:

____

Do you have trouble sleeping or suffer from insomnia? YES □ NO □
If yes, please specify the details:

____

Do you take any medications? YES □ NO □
If yes, please specify which ones:

____

KATILIMCI TIBBİ GEÇMİŞ FORMU

Katılımcı n. __________   **Test tarihi**: __________________

KATILIMCI HAKKINDA GENEL BİLGİLER

Soyadı ve Adı: ____

Doğum Tarihi: ______________________Yaş: _____________________

Eğitim (okul yıl sayısı): ______________________________________

Erkek □ Kadın □   Sağlak □ Solak □ Her iki elini kullanan □

Meslek: ____

(emekliyseniz, emeklilik öncesi mesleğinizi belirtiniz)

SAĞLIK BİLGİLERİ

Görme ve/veya işitme bozukluğunuz var mı? EVET □ HAYIR □

Evetse, lütfen belirtiniz ve düzeltilip düzeltilmediğini yazınız:

____

Herhangi bir sağlık sorununuz var mı? EVET □ HAYIR □

Evetse, lütfen türünü belirtiniz:

____

Hiç inme geçidiniz mi? EVET □ HAYIR □

Evetse, lütfen ayrıntıları belirtiniz:

____

Hiç kafa travması geçirdiniz mi? EVET □ HAYIR □

Evetse, lütfen ayrıntıları belirtiniz:

____

Hiç epileptik nöbet veya kasılma geçirdiniz mi? EVET □ HAYIR □

Evetse, lütfen ayrıntıları belirtiniz:

____

1

Hiç nörolojik danışmanlığa ihtiyaç duydunuz mu? EVET □ HAYIR □
Evetse, lütfen ayrıntıları belirtiniz:

____

Hiç psikolojik veya psikiyatrik danışmanlığa ihtiyaç duydunuz mu? EVET □ HAYIR □
Evetse, lütfen ayrıntıları belirtiniz:

____

Ailenizden herhangi biri psikolojik, psikiyatrik veya nörolojik danışmanlığa ihtiyaç duydu mu? EVET □ HAYIR □
Evetse, lütfen ayrıntıları belirtiniz:

____

Hiç uyuşturucu kullandınız mı ve/veya alkolü kötüye kullandınız mı? EVET □ HAYIR □
Evetse, lütfen ayrıntıları belirtiniz:

____

Uyku problemleri ya da uykusuzluk yaşıyor musunuz? EVET □ HAYIR □
Evetse, lütfen ayrıntıları belirtiniz:

____

Herhangi bir ilaç kullanıyor musunuz? EVET □ HAYIR □
Evetse, lütfen hangileri olduğunu belirtiniz:

____

Section B: MAC-Q in Turkish

Bellek Öz Değerlendirme Anketi	
Geçmişe kıyasla (on yıl önce), aşağıdaki görevlerdeki mevcut yeteneğinizi nasıl tanımlarsınız?	
Yeni tanıştığınız birinin adını hatırlama	Şu anda çok daha iyi Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü
Günlük veya haftalık kullandığınız telefon numaralarını veya adresleri hatırlama	Şu anda çok daha iyi Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü
Evde veya ofiste nesneleri (örneğin anahtarları) koyduğunuz yeri hatırlama	Şu anda çok daha iyi Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü
	Şu anda çok daha iyi

Kısa bir zaman önce okuduğunuz bir gazete veya dergi makalesinden belirli bilgileri hatırlama	Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü
Market veya eczaneye gittiğinizde almayı planladığınız ürünü/ürünleri hatırlama	Şu anda çok daha iyi Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü
Genellikle, lise döneminizdeki hafızanızı şu andaki hafızanıza kıyasla nasıl değerlendirirsiniz?	Şu anda çok daha iyi Şu anda biraz daha iyi Aynı Şu anda biraz daha kötü Şu anda çok daha kötü