

A multi-phase pilot study on multimodal evaluation in cognitive tests

Juan Camilo Méndez Flórez¹, Nelly Condori-Fernández^{1,2,*}, Alejandro Catala^{1,2} and Alejandro Enriquez³

¹Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

²Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

³RadAmbiente, Cuenca, Ecuador

Abstract

Cognitive impairment is an increasingly prevalent issue, particularly with the ageing population. Brief cognitive screening tests play a key role in early detection, and their progressive digitisation raises new methodological challenges. One such test, the SAGE (Self-Administered Gerocognitive Exam), which assesses areas such as orientation, language, and memory, has proven useful for the early detection of cognitive impairment. This paper presents a pilot study designed to compare the traditional and digital formats of a cognitive test based on a multimodal protocol. The protocol integrates objective performance scores, subjective assessments of user experience (UX) and mental workload, and emotional response analysis conducted in two sequential phases: the first using facial expression analysis, and the second focusing exclusively on stress via electrodermal activity (EDA) sensors. The second phase was conducted in care home environments, where ethical constraints prohibited the use of cameras, highlighting the protocol's adaptability to real-world sensitive contexts. This paper contributes to understanding how test format influences not only cognitive outcomes but also UX and perceived mental workload, providing a replicable protocol for evaluating the digital transition of cognitive assessment tools.

Keywords

cognitive impairment, digital SAGE test, pilot study, UX, emotional data, electrodermal activity, workload

1. Introduction

Cognitive impairment is a growing challenge for global public health, driven by an ageing population and aggravated by factors such as unwanted loneliness among older adults [1]. In Spain, this issue is particularly pressing: according to Alzheimer Europe (2019), 1.83% of the population already shows signs of cognitive impairment, a figure projected to more than double by 2050, reaching 3.99% [2]. In this context, the early detection of mild cognitive impairment (MCI) has become a key priority.

Among the tools available for this purpose, the Self-Administered Gerocognitive Exam (SAGE) [3] stands out. This clinically validated, self-administered test enables the early detection of MCI by assessing cognitive domains such as orientation, language, memory, executive function, and visuoconstructive skills [4, 5]. Its ease of use makes it a practical, remote, and scalable option.

The digitisation of the SAGE test has been motivated by the need to increase its accessibility, particularly for vulnerable populations such as older adults with mobility limitations, those living in rural areas, or individuals experiencing social isolation [6]. While mobile or tablet-based versions offer potential benefits such as self-monitoring and immediate feedback, the shift from paper-based to digital formats introduces questions about usability, user experience, and cognitive performance comparability.

This paper presents a pilot study aimed at comparing the traditional and digital formats of a cognitive test, taking SAGE as a representative case. The digital format used in this study corresponds to the version developed in [6]. Our experimental protocol integrates objective performance scores, subjective

Interacción '25: XXV International Conference on Human-Computer Interaction, September 03–05, 2025, Valladolid, Spain

*Corresponding author.

✉ juancamilo.mendez.florez@usc.es (J. C. M. Flórez); n.condori.fernandez@usc.es (N. Condori-Fernández); alejandro.catala@usc.es (A. Catala); alejo.sebas99@outlook.com (A. Enriquez)

ORCID 0009-0006-7282-0835 (J. C. M. Flórez); 0000-0002-1044-3871 (N. Condori-Fernández); 0000-0002-3677-672X (A. Catala)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

assessments of user experience and mental workload, and emotional response analysis conducted in two sequential phases that correspond to different settings. In the first phase, emotional data were gathered via facial expression analysis in a lab setting. In the second phase, conducted in care home environments, emotional responses were monitored through electrodermal activity (EDA) sensors, due to restrictions on video recordings in such settings.

This pilot study highlights the adaptability of a multimodal evaluation to ethically sensitive real-world contexts, where traditional data collection methods may be restricted, offering insights for the development of digital cognitive testing approaches that balance methodological rigour with user acceptability.

The remainder of this paper is organised as follows: Section 2 provides background and reviews related work on the digitisation of cognitive tests. Section 3 presents the methodology, including the study's objective, variables and experimental procedure. Section 4 reports the initial results. Finally, Section 5 discusses some key aspects to be considered in future extended formal evaluations.

2. Background

2.1. Self-administered gerocognitive examination (SAGE) test

Several tools have been developed considering the importance of early detection and prevention. For this study, the Self-Administered Gerocognitive Examination (SAGE) test will be brought into focus. Its reliability and validity as a tool for early detection of dementia and Alzheimer's disease are well-documented through several studies [4]. This test evaluates different cognitive areas, such as:

1. **Orientation:** This domain assesses awareness of time, date, and place, which are essential for daily functioning. Deficits may indicate early cognitive impairment, common in dementia.
2. **Naming:** Assesses the ability to accurately identify and name objects, involving the involvement of the brain's language centers.
3. **Similarities:** Assesses reasoning and abstract thinking by comparing objects or concepts.
4. **Calculation:** It evaluates cognitive skills related to numerical processing, logical thinking, and mathematical problem-solving.
5. **Memory:** This domain assesses the capacity to encode, store, and retrieve new information.
6. **Construction:** Evaluates visuospatial skills and the ability to plan and execute movements for the purpose of copying or constructing figures.
7. **Verbal Fluency:** This area evaluates the ability to generate words rapidly and efficiently under specific constraints, measuring executive function, lexical retrieval, and language fluency.

2.2. Related work

This section critically discusses the similarities and differences between the current study and related literature in early detection and prevention of Alzheimer's and dementia. In terms of the validity of the SAGE test, [5] evaluated 665 patients, comparing the results of the SAGE with the Mini-Mental State Examination (MMSE). The findings revealed that the SAGE detects dementia six months earlier than the MMSE and, being self-administered, removes implementation barriers associated with the need for medical personnel. Additionally, it highlights that repeated SAGE scores function as a reliable cognitive biomarker for monitoring the progression of impairment.

Other experiments have validated the applicability of the test in broader contexts, demonstrating that the SAGE test is feasible, practical, reliable, and effective in various settings and among different population groups [7, 4].

Finally, Scharre et al. [8] found a strong correlation between the paper SAGE scores and its paid digital version, suggesting that the digital format is equally valid. It does not provide details about user experience and usability design issues.

3. Methodology

This section describes the methodological approach of the pilot study. Given its exploratory nature, the study adopts a multi-phase research design in which each phase addresses complementary aspects of the overall goal through specific data collection techniques and settings. This design allowed the protocol to adapt to evolving constraints in both controlled and real-world environments.

3.1. Goal and research questions of the study

The main objective of this study is structured using the Goal–Question–Metric template [9]:

Analyze differences associated with the test format (traditional vs. digital) for the purpose of comparing cognitive performance, perceived user experience, mental workload, and emotional responses with respect to the outcomes observed through a multimodal evaluation approach from the point of view of researchers and practitioners interested in digital cognitive screening in the context of a pilot study conducted in two phases, including real-world ethically sensitive settings.

The study is guided by the following research questions and related variables. The dependent variables measured are detailed in Table 1.

- **RQ1. How different is the cognitive performance between the traditional and digital formats of the cognitive test?**

This question explores whether the format of the cognitive test influences participants' performance, which is crucial for assessing the comparability of both formats in the early detection of mild cognitive impairment (MCI). Cognitive performance is measured using the test scores (range: 0–22) obtained in both the traditional and digital formats of the SAGE test.

- **RQ2. How does the test format (traditional vs. digital) affect the perceived workload?**

This question aims to explore whether participants perceive different levels of workload depending on the test format. In the study, perceived workload was measured using the NASA TLX questionnaire [10], which captures participants' subjective impressions of task demands. The instrument uses a scale to rate six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration.

- **RQ3. How do participants' emotional responses differ depending on the test format?**

This question examines whether participants' emotional responses vary according to the test format. Participants' emotional responses during the completion of the SAGE test were gathered. Emotional data related to frustration (from 0% to 100%, indicating the proportion of time facial expressions related to that emotion were detected during the test), valence (from -1 to 1, indicating negative to positive emotions), and arousal (from -1 to 1, indicating emotional activation levels) were obtained through facial analysis using Kopernica Human, an artificial intelligence tool developed by Neurologyca¹. Additionally, stress was assessed using physiological responses from the EDA sensor, recorded with the EmbracePlus smartwatch device by Empatica². The data were processed by an automatic stress detector that outputs discrete stress levels from 1 to 5, where levels 4 and 5 indicate stress [11].

3.2. Experimental procedure

The study was structured in two phases to evaluate cognitive performance, mental workload, and emotional responses elicited by the SAGE test in both its traditional (paper-based) and digital formats. In both phases, the digital format of the SAGE developed in [6] was used, available as an APK for Android

¹Neurologyca, Kopernica: Real-time emotional intelligence, 2025. URL: <https://www.kopernica.ai/>

²Empatica, EmbracePlus, URL: <https://www.empatica.com/en-eu/embraceplus/>

Table 1

Emotion and workload variables measured, specifying the instruments used and their respective measurement scales.

	Variable	Instrument	Scale
Emotional data	Frustración	Kopernica Human	0-100%
	Valencia	Kopernica Human	-1 to 1
	Arousal	Kopernica Human	-1 to 1
	Stress	Smartwatch EmbracePlus	0.01 to 100 μ S
		Automatic Stress Detector [11]	1 - 5 (stress if level \geq 4)
	Workload	NASA TLX Workload	-10 to 10
	Cognitive Performance	SAGE test	0 to 22

devices and specifically designed for use on touchscreen tablets. Figure 1 shows the experimental procedure and instruments used across both study phases.

The implementation of each phase was subject to ethical approvals and institutional permissions. Participants in both phases were required to meet the following inclusion criteria: no prior formal diagnosis of cognitive impairment, and provision of informed consent.

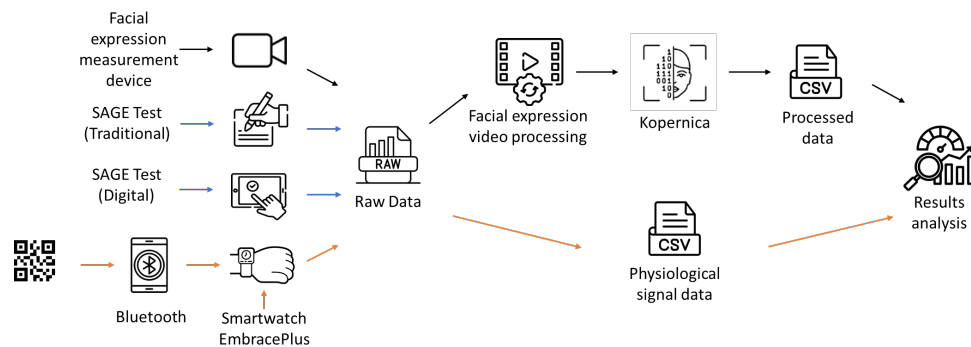


Figure 1: Data Flow and Tools Used Across the Pilot Study Phases: Black arrows indicate Phase 1 (facial expression measurement device and Kopernica for video data), orange arrows indicate Phase 2 (Smartwatch EmbracePlus for physiological data), and the blue arrow indicates both phases (SAGE test).

3.2.1. Phase 1: lab setting

In Phase 1, conducted in a laboratory setting, the SAGE test was administered in both formats under controlled conditions, with a main focus on the evaluation of the digital version. This version was designed based on usability principles for older adults and aimed at reducing cognitive load, minimising errors, and ensuring an accessible experience through large buttons and typography, clear instructions, linear navigation, and a minimalist design with emotionally positive colours (i.e., Blue, White and Yellow).

Participants were selected through convenience sampling, based on their availability, proximity and voluntary participation. To assess emotional responses during the administration of the digital test, participants' facial expressions were recorded using a camera and OBS Studio, a free video recording software configured specifically to capture close-up video in MP4 format. The videos were used as input in the Kopernica tool, which applies facial emotion analysis algorithms and generates CSV files with the corresponding results for interest, frustration, valence, and arousal percentages.

Immediately after the test, participants responded to the NASA TLX Workload questionnaire to assess perceived workload. The order of test format presentation was randomized across participants, and a one-week gap was maintained between the sessions for each format.

3.2.2. Phase 2: home care setting

In Phase 2, carried out in a home care setting, participants were older adults residing in a local nursing home, which allowed for the inclusion of a more diverse population and testing under uncontrolled conditions, in contrast to Phase 1. Each participant completed both the digital and traditional versions of the test, with at least a 7-day interval between each administration to avoid short-term memory effects. In addition to cognitive scores, physiological data were collected during both tests to evaluate user experience.

For physiological data collection, the EmbracePlus smartwatch, a wearable device that records real-time electrodermal activity data through integrated electrodes that must be in contact with the skin to capture accurate information, was used. It operates with a sampling frequency between 1 and 4 Hz, within a range of 0.01 to 100 microsiemens (μS). To initiate monitoring, each participant was previously assigned a unique credential through Empatica's Care Lab Portal, which generates a QR code used to pair the watch with a mobile phone via Bluetooth. Once paired, the device was placed on the non-dominant wrist of the participant. Raw sensor data collected from the wearable device were stored in AVRO format, a compact binary format for efficient data storage and fast access, and subsequently converted to CSV using a Python script provided by Empatica for further analysis. The NASA TLX Workload questionnaire was also administered after each format.

All participants provided informed consent prior to participation, in accordance with the guidelines approved by the relevant ethics committees. In the home care setting, consent from legal guardians or family members, when applicable, was managed directly by the facility staff. To protect participants' privacy, video recording was restricted in Phase 2, and only non-intrusive physiological data were collected using a wearable device. Data from the EmbracePlus smartwatch was managed by Empatica under a research license valid for the duration of the project. For video-based emotional analysis in Phase 1, only numerical output from the tool was stored, and no raw video footage was retained. In line with the agreement established with the care facility, the research team also provided both individual and group-level summaries of test results, along with tailored cognitive health recommendations.

4. Initial results

4.1. Phase 1: lab setting

Five older adults (three individuals aged between 60 and 80 and two individuals over 80) completed the SAGE test in both formats.

Table 2-Phase 1 shows the SAGE test scores for participants in the first phase, comparing both the digital and traditional formats and including the *difference in performance* per participant. The scores showed high consistency between the traditional and digital formats for most participants. For example, P2, P3, and P4 obtained identical scores in both formats (19, 11, and 17, respectively), while the remaining two (P1 and P5) showed a small difference of two points in favour of the traditional version. These results, under controlled conditions, suggest that cognitive performance is generally consistent across formats, although limited digital experience may slightly influence outcomes in favour of the traditional version, likely due to lower familiarity with mobile devices, which is related to age—as in the case of P1, who is 91 years old.

Regarding *mental workload* (see values in Table 3-Phase 1), the NASA TLX Workload data for the digital version showed significant variations among participants. The average workload ranged from -6.0 to 7.2, with P4 reporting the highest workload (7.2) and P2 the lowest (-6.0). Participant P4 indicated a notably demand with the digital format, despite obtaining identical scores in both formats (17). In contrast, P2 reported negative values in most dimensions, suggesting a comfortable and undemanding experience.

In terms of *emotional responses*, the overall analysis of valence, arousal, and frustration (processed through Kopernica) in the digital version revealed that most participants are located in the lower-right quadrant of the affective circumplex, corresponding to a "Calm" emotional state, characterized by

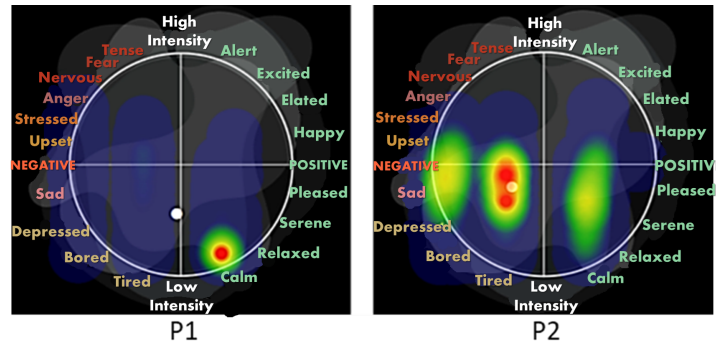


Figure 2: Affective circumplex maps for P1 and P2 based on valence and arousal from facial expression analysis (Kopernica) during the digital test.

positive valence and low arousal (see Figure 2). This suggests that these participants felt generally comfortable and relaxed while completing the test. An exception was observed in the case of Participant P2, who exhibited slightly negative valence and higher arousal, placing between "Sad" and "Upset" zones. According to Figure 3-a, this participant also reported the highest frustration percentage (34%).

These results suggest the digital format is comparable in performance but poses challenges for some participants, reflected in higher frustration and workload. The prevailing calm indicates good acceptance, but cases like P2 highlight the need for more intuitive interfaces to reduce frustration.

Table 2

SAGE test scores by participant, comparing digital and traditional formats across phases

Participant	Traditional Version	App Version	Difference (Trad. – App)
<i>Phase 1</i>			
P1	18	16	2
P2	19	19	0
P3	11	11	0
P4	17	17	0
P5	22	20	2
<i>Phase 2</i>			
U1	8	11	-3
U2	10	17	-7
U3	17	18	-1
U4	16	17	-1
U5	11	14	-3
U6	15	18	-3
U7	14	18	-4

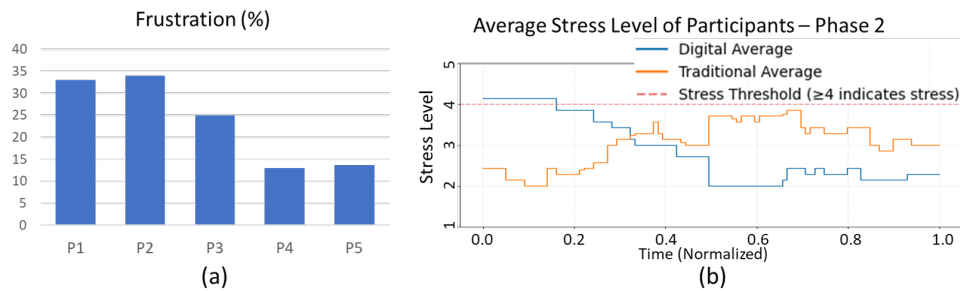
4.2. Phase 2: home care setting

Seven older adults (five women, two men; ages: 90, 76, 75, 80, 70, 90, 80, respectively) completed the SAGE test in both formats in a more realistic setting. Table 2-Phase 2 shows the SAGE test scores for participants in this phase. The group showed a general trend of improvement in the digital format, where average scores increased from 13 (traditional) to 16 (digital). Regarding mental workload (see values in Table 3-Phase 2), some differences between formats were also observed. In the digital format, average workload ranged from -7.7 to 4.5, with U6 showing that the task was demanding despite improvement in performance (14 to 18). U4, with the lowest average (-7.7), reported negative values across all dimensions, indicating a very comfortable experience. In the traditional format, averages were more homogeneous, ranging from -3.7 to 1.0. However, the overall trend indicates that while the digital format improved performance, it may introduce additional demands for some.

Table 3

Perceived workload scores by participant for digital and traditional formats across both phases

Participant	Mental Demand		Physical Demand		Temporal Demand		Performance		Effort		Frustration		Avg. App	Avg. Trad.
	App	Trad.	App	Trad.	App	Trad.	App	Trad.	App	Trad.	App	Trad.		
Phase 1														
P1	3	–	0	–	0	–	–6	–	–10	–	–10	–	–3.8	–
P2	–10	–	–10	–	–10	–	0	–	–6	–	0	–	–6.0	–
P3	0	–	2	–	0	–	0	–	3	–	3	–	1.3	–
P4	7	–	6	–	6	–	7	–	7	–	10	–	7.2	–
P5	–7	–	–9	–	–4	–	–4	–	–4	–	–6	–	–5.7	–
Phase 2														
U1	–6	–6	0	–10	0	0	2	10	–4	–4	–8	–10	–2.7	–3.3
U2	0	0	0	–6	0	0	5	–4	0	6	0	4	0.8	0.0
U3	–10	0	–10	–10	0	2	–10	–8	–6	4	–10	–10	–7.7	–3.7
U4	10	6	6	–10	6	0	7	10	7	0	–9	–10	4.5	–0.7
U5	–8	4	–7	2	3	–4	–7	4	0	0	–6	0	–4.2	1.0
U6	–6	6	–6	–4	0	4	5	10	2	–4	–7	–6	–2.0	1.0
U7	0	0	0	–6	0	0	–10	–8	0	0	0	2	–1.7	–2.0

**Figure 3:** (a) Frustration during the digital test (Phase 1) from facial expressions. (b) Average stress in Phase 2 from the automatic stress detector

The Stress levels graphs, (see Figure 3-b), revealed that stress was generally lower in the digital format, with lower average peaks compared to the traditional version. These results indicate that in this group of individuals, overall cognitive performance improved in the digital format and stress was reduced, but variations in workload were introduced.

Table 4 summarises stress-related measures for each participant and test format in Phase 2. The table includes the percentage of time spent in high stress levels (≥ 4), average stress level, standard deviation, and the number of detected stress episodes based on EDA signals. When comparing formats, most participants exhibited slightly higher stress levels and more frequent stress episodes in the traditional version than in the digital one. This pattern is particularly evident in the percentage of time spent at stress levels 4 and 5, which tended to be greater under the paper-based condition. These differences may be partially explained by the improvements made to the digital prototype following usability feedback in Phase 1. Additionally, the fixed font size and layout of the paper-based test could have introduced visual strain for some participants.

5. Discussion and conclusions

This pilot study tested a multimodal protocol in two distinct settings to compare cognitive performance, workload, and emotional responses across digital and traditional formats of a cognitive test. Several aspects of the protocol proved effective across both phases. Paper-based instruments—used for the traditional cognitive test and workload measurement—were reliable and well accepted in both contexts. The informed consent process was also simple, with most participants signing voluntarily without raising concerns.

In both settings, certain test items required verbal clarification during administration, and a think-aloud protocol could have enriched feedback beyond written responses. Most participants completed both test formats, with two exceptions in Phase 2 due to contextual constraints.

Scheduling was more consistent in Phase 1, with randomised order and a one-week interval between

Table 4
Stress indicators by participant and test format (phase 2)

Participant	Test Format	% Time (at levels 4 and 5)	Mean Stress Level	Standard Deviation	Stress Episodes
U1	Traditional	50	3.1	1.5	3
	Digital	16.7	2.8	1.1	1
U2	Traditional	71.4	3.3	1.5	3
	Digital	25	2.8	1.5	1
U3	Traditional	40	3.0	1.4	2
	Digital	33	3.0	1.0	2
U4	Traditional	60	3.2	1.5	3
	Digital	33	3.0	1.3	2
U5	Traditional	33	2.8	1.7	1
	Digital	43	2.9	1.6	2
U6	Traditional	50	3.1	1.2	4
	Digital	50	3.2	1.6	2
U7	Traditional	17	2.8	1.0	2
	Digital	20	2.5	1.0	2

formats. In Phase 2, scheduling varied due to participants' availability or health issues, extending in some cases to three weeks or resulting in partial data.

Distinct challenges emerged in each phase. In Phase 1, one of the main difficulties involved managing and analysing video recordings used for emotion detection via Kopernica. The videos had to be processed offline, and in several cases, facial detection accuracy was compromised due to posture. Therefore, emotion recognition based on facial expressions is only available for the digital format. A few participants also expressed concern about receiving feedback on their cognitive performance, which may reflect a general sensitivity to test results in this demographic.

In Phase 2, ethical restrictions within the home care setting strongly influenced data collection. Video recording was not allowed at all; instead, physiological data were gathered through the EmbracePlus wearable as an acceptable alternative. Although this method limited emotional data granularity, restricting it mainly to EDA-based stress levels, it was well tolerated. Figure 3-b shows that stress in digital format starts high but decreases and stabilises, while on paper it gradually increases. Table 4 confirms that most participants experienced more episodes and longer durations of elevated stress on paper, possibly due to usability improvements in the digital version and visual discomfort with the traditional format. Participants often reported being unaware of the device. Thus, using non-obstructive acquisition methods could be a preferable alternative in settings where administrators/users are reluctant to camera-based methods.

Participant recruitment in Phase 2 was coordinated by the facility, which limited researchers' ability to select or balance participant profiles. Still, the core testing instruments were applied consistently across both phases, with variation only in the method of emotion measurement. These contextual and methodological differences were documented and considered in interpreting results.

Overall, this paper contributes a protocol for conducting a pilot involving two different types of settings, which aimed at gauging differences in cognitive performance and workload influenced by the implementation format of the SAGE tests. The pilot reported initial results, providing some encouraging preliminary results on whether the digital version may pose or not a serious entry barrier before running a more formal experiment. It also illustrated how multimodal measurements concerned with emotional responses can be used to complement the performance and workload measurements.

This pilot study highlights the importance of planning for granular physiological data collection, balanced test scheduling, and adaptive digital design when working with older adults. Despite contextual

constraints, the proposed multimodal evaluation protocol proved adaptable to both controlled and real-world settings. In this regard, the consistent use of the core instruments across both phases, despite differences in emotional measurement, supports the protocol's scalability. However, ethical aspects such as informed consent, data privacy, and participant vulnerability must be adapted to each setting. Addressing them early is key to ensuring ethical and contextual suitability at scale.

These observations support the feasibility of replicating this multi-phase approach in similar studies evaluating digital transitions in cognitive testing. The protocol offers a comprehensive framework to examine the impact of test format in elderly populations.

Acknowledgements

Authors acknowledge the support of the Galician Ministry of Culture, Education, Professional Training and University (grants ED431G2023/04 and ED431C2022/19). Supported also by Interreg VI-A Spain-Portugal Program (POCTEP) 2021-2027 with grant 0144_TRANSFIRESAUDE_1_E, CNS2024-154915 by MCIN/AEI, the Erasmus Mundus Joint Master Degree program SE4GD-619839, and the ERDF.

Declaration on generative AI

During the preparation of this work, the author(s) used Grammarly for Grammar and spelling checks.

References

- [1] L. Qiao, G. Wang, Z. Tang, S. Zhou, J. Min, M. Yin, M. Li, Association between loneliness and dementia risk: A slr and meta-analysis of cohort studies, *Front. Hum. Neurosci.* 16 (2022).
- [2] Alzheimer Europe, Dementia in Europe Yearbook 2019: Estimating the prevalence of dementia in Europe, Alzheimer Europe, 2019.
- [3] Wexner Medical Center, Sage: The self-administered gerocognitive exam, 2023.
- [4] D. W. Scharre, S.-I. Chang, R. A. Murden, J. Lamb, D. Q. Beversdorf, M. Kataki, H. N. Nagaraja, R. A. Bornstein, Self-administered gerocognitive examination (sage): A brief cognitive assessment instrument for mild cognitive impairment (mci) and early dementia, *Alzheimer Disease Associated Disorders* 24 (2010) 64–71.
- [5] D. W. Scharre, S.-I. Chang, H. N. Nagaraja, N. C. Wheeler, M. Kataki, Self-administered gerocognitive examination: longitudinal cohort testing for the early detection of dementia conversion, *Alzheimers Res. Ther.* 13 (2021) 192.
- [6] A. Enriquez Mancheno, Digitization of the SAGE test: building a scalable solution for early cognitive impairment detection, Master's thesis, Lappeenranta–Lahti University of Technology LUT, Finland, 2024.
- [7] D. W. Scharre, S.-I. Chang, H. N. Nagaraja, J. Yager-Schweller, R. A. Murden, Community cognitive screening using the self-administered gerocognitive examination (sage), *J. Neuropsychiatry Clin. Neurosci.* 26 (2014) 369–375.
- [8] D. W. Scharre, S. I. Chang, H. N. Nagaraja, N. E. Vrettos, R. A. Bornstein, Digitally translated self-administered gerocognitive examination (esage): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments, *Alzheimers Res. Ther.* 9 (2017) 44.
- [9] V. R. Basili, G. Caldiera, H. D. Rombach, The goal question metric approach, in: J. J. Marciniak (Ed.), *Encyclopedia of Software Engineering*, John Wiley & Sons, NY, USA, 1994, pp. 528–532.
- [10] Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: P. A. Hancock, N. Meshkati (Eds.), *Human Mental Workload*, volume 52 of *Adv. Psychol.*, North-Holland, 1988, pp. 139–183.
- [11] F. S. Lopez, N. Condori-Fernández, A. Catalá, Towards real-time automatic stress detection for office workplaces, in: J. A. Lossio-Ventura, D. Muñante, H. Alatrística-Salas (Eds.), *Information Management and Big Data*, volume 898 of *Commun. Comput. Inf. Sci.*, Springer, 2018, pp. 273–288.