Social robots as psychometric tools for cognitive assessment: a pilot test

VARRASI, Simone, DI NUOVO, Santo, CONTI, Daniela <http://orcid.org/0000-0001-5308-7961> and DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>

Social robots as psychometric tools for cognitive assessment: a pilot test

Simone Varrasi [1], Santo Di Nuovo [1], Daniela Conti [2], Alessandro Di Nuovo [2]

[1] Department of Educational Sciences, University of Catania, Italy
[2] Sheffield Robotics, Sheffield Hallam University, UK

Abstract. Recent research demonstrated the benefits of employing robots as therapeutic assistants and caregivers, but very little is known on the use of robots as a tool for psychological assessment. Socially capable robots can provide many advantages to diagnostic practice: engage people, guarantee standardized administration and assessor neutrality, perform automatic recording of subject behaviors for further analysis by practitioners. In this paper, we present a pilot study on testing people's cognitive functioning via social interaction with a humanoid robot. To this end, we programmed a social robot to administer a psychometric tool for detecting Mild Cognitive Impairment, a risk factor for dementia, implementing the first prototype of robotic assistant for mass screening of elderly population. Finally, we present a pilot test of the robotic procedure with healthy adults that show promising results of the robotic test, also compared to its traditional paper version.

Keywords: Assistive Robotics; Social Robot; Cognitive Assessment; Human-Robot Interaction; Dementia screening.

1 Introduction

The use of robots as therapeutic assistants and caregivers is one of the most investigated application of robotics in clinical and health psychology. Indeed, the efficacy of artificial agents has been documented with children [1, 2] as well as with elderly people in a variety of neurological and psychiatric conditions [3, 4], leading scholars to suggest a stable integration of Human-Robot Interaction (HRI) in healthcare [5]. Socially Assistive Robotics is the field where most of the research has been focusing since its definition was provided [6], but service robotics and smart environments have been extensively explored too, especially for the elderly who may need complex assistive technology to support healthy ageing [7, 8]. Robotics can also assist with the cognitive rehabilitation [9] and to build models of cognitive dysfunctions [10].

It is evident that robots play a promising role in mental health field. However, there are still many unknown or poorly understood fields of application. One of these is the psychological assessment.

According to Scassellati [11], a great hope for robotics in Autistic Spectrum Disorder (ASD) is also the implementation of objective measurement of social behavior. This idea is encouraged by the fact ASD manifests behaviorally and diagnosis comes

from the observation of developmental history and social skills, and clinicians do not always agree when evaluating the same patient [12]. Despite these solid motivations, few prototypes of ASD robotic evaluations are described in the scientific literature. Petric [13] has tried to structure a "robotic autism spectrum disorder diagnostic protocol", in order to evaluate child's reaction when called by name, his/her symbolic and functional imitative behavior, his/her joint attention, and his/her ability to communicate via multiple channels simultaneously, but the results are not clear and definitive [14]. Another diagnostic method for ASD is proposed by Wijayasinghe et al. [15], who see in HRI a way to objectively evaluate imitation deficits. In this case, the robot Zeno performs upper body gestures and the child should imitate them, while the robot automatically assesses the child's behavior.

However, it seems not much work has been done regarding other pathologies. Kojima et al. [16], for instance, published some speech recognition experiments with elderly people using the robot PaPeRo, aimed at the development of a computerized cognitive assessment system.

Therefore, at the best of our knowledge, robotic psychological assessment is almost unexplored. The evidence is limited, the pathologies studied are very few, and comparative studies (robotic assessment vs human assessment; robotic assessment vs computerized assessment) are not available. It is evident that more research and experimental evaluation are strongly needed.

From our point of view, the advantages of using a robotic assessor would be multiple: time-saving, quick and easy updates, widely available tools, standardization, the avoidance of assessor bias, the possibility of micro-longitudinal evaluations, scoring objectivity, and having a recording of the administration. Robots can be programmed to perform specific actions always in the same way, so standardization is one of their most interesting features. Therefore, the robotic implementation of quick screening tests could be promising, because they are often repetitive and easy to take, but time-consuming for staff. A robot could administer them, automatically score them, and transmit the result to the psychologist, who could then decide whether to continue with other human-performed tests or not. In fact, robots must not formulate diagnoses, but would provide preliminary diagnostic information about patients to reduce the workload for humans and increase the population that can be screened.

To make a step toward this direction, this paper presents results of a pilot study whose aim was to implement a screening tool for Mild Cognitive Impairment (MCI) [17] on a social robot. MCI used to be considered merely as a prodromal stage of dementia, but today it is recognized as a risk factor to develop more severe cognitive deterioration [18, 19]. It is very important to detect the so-called predictors of conversion and to determine if a patient may develop dementia for efficient planning of a prompt intervention with adequate treatment. For early detection, the markers to be considered can be both biological and psychometric [20]. However, the role of psychometric tests is crucial, because they are quicker and inexpensive, indeed cognitive deficits are usually diagnosed first this way.

This pilot study presented here had three main goals: 1) to develop a first robotic version of an MCI-specific test, fully administered and scored by a social robot; 2) to

compare robotic test administration to traditional paper administration; 3) to collect data and information for further improvements.

2 Materials and Methods

2.1 Participants

As the research was not meant to provide a first validation on a clinical sample, but rather a preliminary proof of the viability of the robotic psychometric approach, we chose to enroll healthy adults (n = 16, Males = 10, Females = 6, M-age = 31.5 years, range = 19-61, SD = 14.15) among university staff and students. We selected people who had lived and worked in the UK for at least four months (M = 110.19, SD = 188.23) and we recorded their years of education as well (M = 19.5, SD = 4.07).

2.2 The MoCA test

The Montreal Cognitive Assessment, better known as the MoCA test, is a brief cognitive screening tool for Mild Cognitive Impairment [21] freely available from the official website, used in 100 countries around the world and translated into 46 languages. It is composed of eight subtests: visuospatial/executive (alternating trail making, copying a cube, drawing of the clock), naming, memory, attention (digit span, vigilance, serial 7s), language (sentence repetition, fluency), abstraction, delayed recall, and orientation. The maximum score is 30, and a score equal to 26 or above is considered normal. If the person has twelve years of education or less, one point is added. In this project, the Full 7.1 English version inspired the implementation, leading to a new robotic screening test for MCI. The English version of the MoCA test has two alternative versions, 7.2 and 7.3, equivalent to the main Full 7.1 version [22]. The 7.2 version was used in this study for comparison.

2.3 The Pepper robot and the cognitive test software implementation

Our robotic cognitive test assesses the same areas of the MoCAs, therefore we have the same subtests that for simplicity have the same names.

The platform used in our experiments is the humanoid Pepper, the latest Aldebaran's and SoftBank Robotics' commercial product specifically designed for HRI and equipped with state-of-the-art interactive interfaces: touchscreen, human-like movement, pressure sensors, object recognition, speech production and comprehension, age, gender, emotion and face detector. Pepper supports different programming languages, such as Python, Java, Silverlight and C++ SDK.

To implement the prototype used in this work we used the Choregraphe suite (version 2.5.5), that provides a drag and drop interface that allows building algorithms and the robot’s behaviors starting from pre-existing boxes with accessible Python code, which was modified to suit the needs of the particular implementation used in this work.

2.4 Experimental Procedure

The administration instructions reported in the English MoCA manual inspired the implementation, and two more tasks, one at the beginning of the test (the welcome task) and one at the end (the thank you task) were added. In the welcome task, Pepper introduced itself and asked the participant to provide his/her age, gender and years of education. This was meant both to collect important information about the person, as well as to train him/her on HRI. The robot could recognize and follow the face in front of it for to better engaging the participant in the interaction [23], and it moved its arms and hands as suggested in the literature in the case of HRI with adults [24].

The final result was a new psychometric test fully administered and scored by Pepper. The language of the administration was English, and the voice used was the robotic one already available in the tool. The administration was standardized, so Pepper always performed the same way, and did not change according to the participant's reactions, and it repeated the instructions only when allowed to do so by the manual. The timing of the administration was regulated by internal timers that were set empirically. Therefore, if the participant did not complete a task, the session continued when those internal timers expired. Pepper audio-recorded the whole session and took photos of the second and third tasks' drawings; moreover, it produced a Dialog file with the transcription of the verbal conversation with the participant and a Log file containing information about any technical failures that occurred, the automatic score achieved, any wrong answers received and any tasks ignored. This way, a clinical psychologist could fully review the administration and re-evaluate it if needed.

The administration phase was divided into two sessions for each participant: the robotic administration and the traditional paper administration. The participants were invited to enter one by one and asked, first of all, to read and sign the forms regarding the processing of personal data.

The robotic session (Figure 1) was entirely run by Pepper: it gave the instructions, registered the answers and calculated the scoring. The experimenter did not interfere with the interaction and maintained a marginal position. The session was video-recorded and timed. After, each participant provided feedback and comments about the experience.

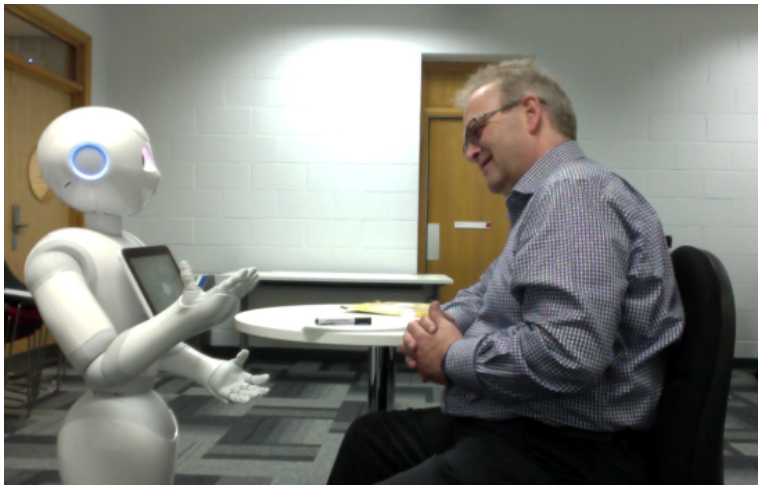

Figure 1. Example of the human-robot interaction during the robotic administration

The traditional paper session was entirely run by the experimenter and timed as well. It is important to note that we balanced the order of the administrations by creating two subgroups: one experienced the robotic administration first, and then the traditional paper administration, while the other experienced the reverse order. Moreover, since we wanted to avoid any learning effect, for each participant the two sessions were spaced by at least five days, and the 7.2 alternative validated version of the MoCA test was used in the traditional paper session.

2.5 Data analysis

Data collected in our pilot experiment were analyzed via classical statistical methods performed with the SPSS software (version 24).

For each participant in our experiment, we derived three scores using different evaluation modalities: 1) Standard score, which is the result of the paper and pencil administration of the test calculated using the MoCA's manual; 2) Automatic score, which is the result of the electronic test administered and automatically calculated by the robot using the build in software; and 3) Supervised score, which is the score calculated by a psychologist, who corrected the automatic score via the video and audio analysis. In our vision, modalities 2 and 3 are a simulation of the procedure for the actual application of a robotic assessment. The automatic score will be used for large population screening, while the supervised score will be calculated by a psychologist for a subset of subjects who are indicated by the automatic scoring as below the threshold and in need of a deeper analysis.

The result section presents the descriptive statistics of the three scores modalities for each subtest and the global score: minimum (Min), maximum (Max), mean (*M*) and standard deviation (*SD*). The correlations among the global scores and the subtest scores (Spearman or Pearson according to the nature of the data and the shape of the distribution) are evaluated to analyze if robot administration scores can fit the stand-

ard score. This aims to assess the concurrent external validity of the robotic procedure by comparing the automatic score with an external criterion, i.e. the MoCA score. Spearman-Brown Coefficient and Cronbach alpha are also calculated to analyze the reliability of automatic scoring, compared with the other modalities, and a regression analysis identifies a predictive function that can relate the automatic score to the standard score and allows deriving the latter from the former.

3 Experimental Results

3.1 Global analysis

The results report that mean global automatic score is 12.69 (Min = 6.00; Max = 23.00; *SD* = 4.61), mean global supervised score is 18.62 (Min = 10.00; Max =27.00; *SD* = 4.83) and mean global standard score is 25.00 (Min = 21.00; Max = 28.00; *SD* = 2.07), as shown in Figure 2.

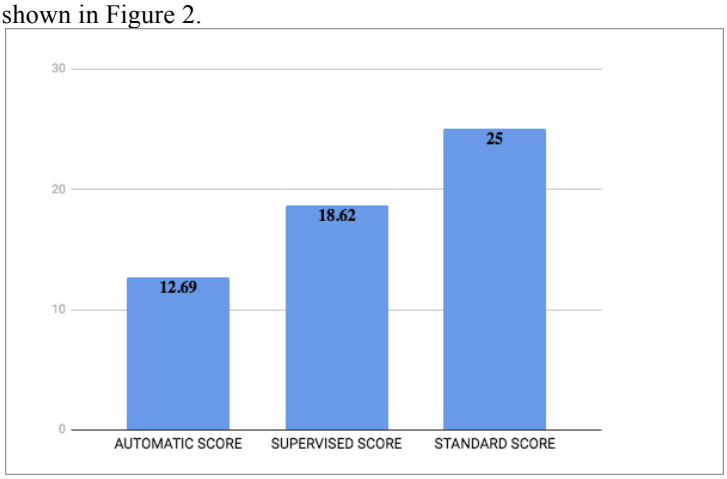

Figure 2. Mean global scores

3.2 Subtest analysis

The automatic score. In the automatic score version, some subtests do not even reach half the maximum achievable score: visuospatial/executive (0.25/5), language (0.63/3) and attention (1.31/6). The other are quite acceptable: naming = 1.94/3, abstraction = 1.06/2, delayed recall = 2.69/5, orientation = 4.44/6 (Figure 3).

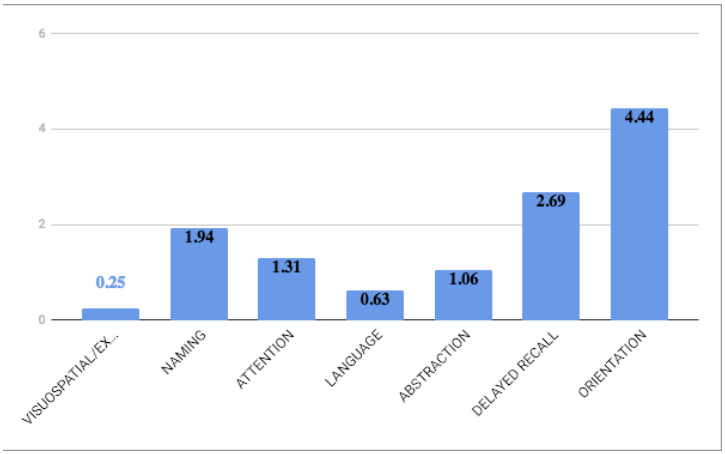

Figure 3. Subtests of the automatic score

The supervised score. In the supervised score version, the subtests that do not reach half the maximum achievable score are abstraction (0.50/2), language (0.75/3) and visuospatial/executive (2.13/5). The other are as follows: naming = 2.81/3, attention = 4/6, delayed recall = 3.13/5, orientation = 5.31/6 (Figure 4).

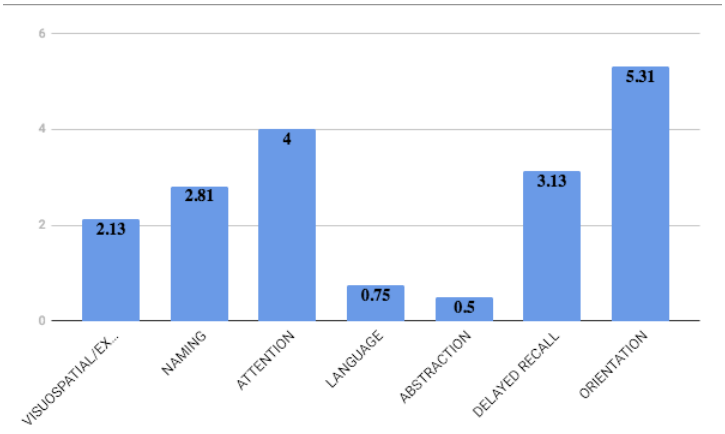

Figure 4. Subtests of the supervised score

The standard score. In the standard score version, all the subtests reach at least half the maximum achievable score (visuospatial/executive = 4.5/5, naming = 2.81/3, attention = 4.56/6, delayed recall = 4.31/5, orientation = 5.94/6), but abstraction (1.13/2) and language (1.75/3) are lower than the others (Figure 5).

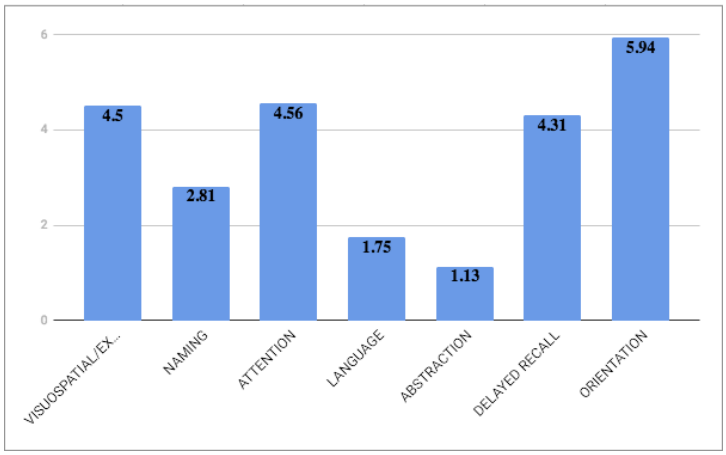

Figure 5. Subtests of the standard score

3.3 Spearman correlations between the global scores

Spearman rank-order correlations, more suitable for the shape of distribution preliminarily assessed for these data, were calculated between the global scores (Table 1). There is a strong relationship between the supervised score and the standard score ($\rho = 0.64$), it is statistically significant ($p < .01$) and its strength is over the high effect-size (> 0.50) according to Cohen's [25] criteria. This confirms that the robotic procedure has a promising validity. However, the correlation between the automatic and standard scores is not significant both from a statistical point of view, as well as from an effect-size point of view ($\rho = 0.01$). The correlation between the automatic and the supervised scores ($\rho = 0.38$) can be considered an indicator of inter-rater reliability and confirms, from an effect-size point of view, the fact that the supervised score is a fixed version of the automatic score.

SCORE	AUTOMATIC	SUPERVISED	STANDARD
AUTOMATIC	*1*		
SUPERVISED	*0.38*	*1*	
STANDARD	*0.01*	***0.64****	*1*

Table 1. Spearman correlations between the global scores (* $p < .01$)

3.4 Pearson correlations between the subtests

Pearson correlations between the corresponding subtests of the three scoring versions were calculated. In the following tables, we used these abbreviations for the subtest names: *V/E* for visuospatial/executive, *Nam* for naming, *Att* for attention, *Lan* for language, *Abst* for abstraction, *D/R* for delayed recall and *Orie* for orientation.

Automatic score subtests vs supervised score subtests. Majority automatically-scored subtests strongly and significantly correlate with the corresponding subtests of the supervised version, as shown in Table 2. It is interesting to note that automatically

calculated language score also correlates with abstraction ($r = 0.67$, $p < .01$) and delayed recall ($r = 0.62$, $p < .01$) calculated in the supervised manner. This suggests that language may be involved in conceptual categorization and encoding/retrieval processes, providing an explanation for the low mean scores of the abstraction and language subtests in the standard version. If confirmed, this finding would justify the standard version's low mean global score as a problem of sampling, considering that participants were predominantly non-native speakers. The other correlations were not significant and below the medium effect-size.

		Supervised score						
		V/E	**NAM**	**ATT**	**LAN**	**ABST**	**D/R**	**ORIE**
Automatic score	**V/E**	***0.61****						
	NAM		*0.14*					
	ATT			***0.53****				
	LAN				***0.76****	***0.67****	***0.62****	
	ABST					*0.25*		
	D/R						***0.73****	
	ORIE							*0.18*

Table 2. Automatic vs supervised subtest correlations (* $p < .05$; ** $p < .01$)

Automatic score subtests vs standard score subtests. The only significant and strong correlation is between abstraction and language ($r = 0.61$, $p < .01$), confirming the finding discussed above. The other correlations are below the medium effect-size (except naming: $r = 0.32$) and not significant, as shown in Table 3.

		Standard score						
		V/E	**NAM**	**ATT**	**LAN**	**ABST**	**D/R**	**ORIE**
Automatic score	**V/E**	*0.20*						
	NAM		*0.32*					
	ATT			*0.29*				
	LAN				*0.28*			
	ABST				***0.61****	*-0.21*		
	D/R						*0.02*	
	ORIE							*-0.14*

Table 3. Automatic vs standard subtest correlations (** $p < .01$)

Supervised score subtests vs standard score subtests. The attention subtest is the only one that shows a significant correlation with its corresponding subtest in the other version ($r = 0.69$, $p < .01$), while the others are not significant and below the medium effect-size, except for delayed recall vs delayed recall ($r = 0.31$). The orientation subtest, instead, correlates strongly and significantly with naming ($r = 0.63$, $p < .01$) and delayed recall ($r = 0.69$, $p < .01$), as shown in Table 4.

		Standard score						
		V/E	**NAM**	**ATT**	**LAN**	**ABST**	**D/R**	**ORIE**
Supervised score	**V/E**	*0.20*						
	NAM		*0.18*					
	ATT			***0.69****				
	LAN				*0*			
	ABST					*-0.13*		
	D/R						*0.31*	
	ORIE		***0.63****				***0.69****	*0.06*

Table 4. Supervised vs standard subtest correlations (** $p < .01$)

3.5 Reliability of the automatic score

In order to check the reliability of the automatic score, the Spearman-Brown Coefficient and the Alpha Coefficient were calculated and found to be 0.73 and 0.67, respectively. According to the Spearman-Brown Coefficient, the automatic score shows a moderate split-half reliability. The Alpha Coefficient, then, shows an internal consistency just under the acceptable minimum (alpha < 0.70). Considering the strong limitations of the automatic scoring system that will be further discussed, it is quite surprising to find such high-reliability scores. We can justify them with the single items' mean scores, which were homogeneously low.

3.6 Multiple linear regression

Multiple linear regression was performed in order to find the subtests that affected the automatic and standard scores the most. The model explains most of the dependent variable's variance (Multiple R = 0.99; Multiple R-Squared = 0.98). The attention and delayed recall subtests affect both the automatic score and the standard score (Table 5) the most. The other independent variables significantly influence the two scores as well, but with a different percentage of variance explained.

AUTOMATIC SCORE			**STANDARD SCORE**		
Effect	*Std. coeff.*	*p*	*Effect*	*Std. coeff.*	*p*
Attention	*0.40*	< .01	Attention	*0.79*	< .01
D/R	*0.27*	< .01	D/R	*0.52*	< .01
Abstraction	*0.26*	< .01	Language	*0.41*	< .01
Language	*0.21*	< .01	V/E	*0.35*	< .01
Orientation	*0.20*	< .01	Abstraction	*0.35*	< .01
Naming	*0.19*	< .01	Naming	*0.20*	< .01
V/E	*0.10*	< .05	Orientation	*0.12*	< .01

Table 5. Subtests' influence on the variance of the automatic and standard scores

4 Discussion

The issues occurred during the robotic administration were both automatic scoring system errors and HRI errors. They will be briefly noted, and the number of participants affected by them will be indicated in brackets.

The automatic scoring system did not assign the point if the participant repeated the digit span sequence slowly (9/16), even if it was correct; it also assigned only 0 or 3 points in the serial 7s task, because it only recognized the five correct subtractions, while the manual allows for assigning intermediate points in case of errors (2/16). In the visuospatial/executive task, Pepper had to recognize the drawn cube and clocks with its object recognition function, but there was a high probability that the object recognition function failed, because more samples of cubes and clocks were needed to teach Pepper all the possible correct versions of drawings (16/16). Moreover, in the drawing of the clock it was possible to assign only 0 or 3 points, because the object recognition function was not sensitive enough to separately evaluate contour, numbers and hands (16/16).

Then, HRI was affected by errors as well, such as unclear pronunciation of the robot, crashes of the interfaces, low usability, lack of intuitiveness, and so on. This happened, for instance, in alternating trail making (crash of tablet: 10/16), vigilance (unclear pronunciation of the instructions: 10/16), copy of the cube (not enough time to draw: 2/16; positioning of the sheet in front of wrong sensors: 2/16; no positioning of the sheet at all: 2/16) and drawing of the clock (unclear pronunciation of the instructions: 6/16; not enough time to draw: 1/16; positioning of the sheet in front of the wrong sensors: 1/16).

Finally, it is important to note that even the standard version of the test was affected by errors, in particular, the tasks of language and abstraction showed lower mean scores than others. This finding can be explained by cultural bias because most of the participants were not native speakers. Moreover, it is important to underline that language and cultural issues may have affected all the administration, by complicating the comprehension of each task.

However, the robotic administration shows some positive aspects too. The welcome, naming, memory, fluency, abstraction and delayed recall tasks worked well, and require little and quick adjustments. Then, participants reacted positively to Pepper and they judged it as “friendly” and “cute”. According to one participant, after some initial diffidence, the robot turned out to be better than PC and tablets, because HRI was “more dynamic” and “more engaging”. The external concurrent validity of robotic administration is promising, even if this comes out from the correlation between the supervised score and the standard score only. This means that automatic scoring errors are the first that have to be fixed in order to obtain a better validity of the procedure, which will be further improved by the correction of HRI issues too. Even reliability is interesting and nearly acceptable, and multiple linear regression shows a good initial fit between the robotic and the standard version of the test.

5 Conclusion and Future work

In this paper, we presented a pilot study on the use of a robotic platform for the administration and support of the scoring of a MoCA-inspired psychometric test, which is widely used for the diagnosis of Mild Cognitive Impairment. The size of the sample is small, but overall results are promising and represent a first step towards the in-depth comprehension of artificial agents' contribution to psychological assessment, even if there are many aspects that should be further investigated.

First of all, automatic scoring is prone to errors, because of the current limitation of the technology and the HRI interfaces that require many refinement cycles before being fully reliable. For example, we found that Pepper's voice should be improved to make it clearer and less childish, for instance by using a recorded human voice. Then, answer modality should allow redundant and interchangeable multimodal interfaces, so that the person can choose how to interact (e.g. speaking or touching) in case of both personal preference and technical issue.

The automatic scoring should be adequately verified and tested before actual use of any psychometric instrument. Furthermore, error causes should be investigated in detail in order to program the robotic system to flag the case for further investigation by a qualified human psychologist.

Therefore, in future work we aim to perform a test with a larger and balanced sample of native speakers, in order to perform a more accurate comparison and to investigate the effect of age and other variables on HRI. Finally, the validation on a clinical sample will be the following step if the validity and usefulness are confirmed.

Apart from the future work discussed above, there are other themes opened by this pilot study. For instance, can robots influence diagnoses? Do they affect the perception of setting and psychological assessment? Furthermore, one of the common psychometric problems - particularly in forensic settings - is the management of deception and malingering. The patient may deceive the robot and bias the test results, for example by writing down the words that s/he should recall, and the robot may not be able to notice this. The solution should come from the use of robots in controlled environments only and from video and audio recordings of the administration. However, this may lead to concerns about privacy and other ethical issues, which need to be examined in more detail.

Even if this field is completely new, we think that a robotic aid in the first phase of diagnostic path would be useful, not only to detect those that already needs clinical assistance but also to provide automatic large screening exams for prevention.

Acknowledgment

The authors gratefully thank all university staff and students who participated in this study. The work was supported by the European Union's H2020 research and innovation program under the MSCA-Individual Fellowship grant agreement no. 703489.

References

1. Conti, D., Di Nuovo, S., Trubia, G., Buono, S., Di Nuovo, A.: Use of Robotics to Stimulate Imitation in Children with Autism Spectrum Disorder: A Pilot Study in a Clinical Setting. In: Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN. pp. 1–6 (2015).
2. Conti, D., Di Nuovo, S., Buono, S., Di Nuovo, A.: Robots in education and care of children with developmental disabilities : A study on acceptance by experienced and future professionals. Int. J. Soc. Robot. 9, 51–62 (2017).
3. Rabbitt, S.M., Kazdin, A.E., Scassellati, B.: Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clin. Psychol. Rev. 35, 35–46 (2015).
4. Conti, D., Cattani, A., Di Nuovo, S., Di Nuovo, A.: A Cross-Cultural Study of Acceptance and Use of Robotics by Future Psychology Practitioners. In: Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN. pp. 555–560 (2015).
5. Iroju, O., Ojerinde, O., Ikono, R.: State Of The Art : A Study of Human-Robot Interaction in Healthcare. I.J. Inf. Eng. Electron. Bus. 9–3, 43–55 (2017).
6. Feil-Seifer, D., Matarić, M.J.: Defining socially assistive robotics. In: Proceedings of the 2005 IEEE 9th International Conference on Rehabilitation Robotics. pp. 465–468 (2005).
7. Di Nuovo, A., Broz, F., Cavallo, F., Dario, P.: New Frontiers of Service Robotics for Active and Healthy Ageing. Int. J. Soc. Robot. 8, 353–354 (2016).
8. Di Nuovo, A., Broz, F., Wang, N., Belpaeme, T., Cangelosi, A., Jones, R., Esposito, R., Cavallo, F., Dario, P.: The multi-modal interface of Robot-Era multi-robot services tailored for the elderly. Intell. Serv. Robot. (2018).
9. Matarić, M.J., Scassellati, B.: Socially Assistive Robotics. In: Siciliano, B. and Khatib, O. (eds.) Springer Handbook of Robotics. pp. 1973–1994. Springer International Publishing, Cham (2016).
10. Conti, D., Di Nuovo, S., Cangelosi, A., Di Nuovo, A.: Lateral specialization in unilateral spatial neglect: a cognitive robotics model. Cogn. Process. 17, 321–328 (2016).
11. Scassellati, B.: How social robots will help us to diagnose, treat, and understand autism. In: Robotics research. pp. 552–563 (2007).
12. Scassellati, B., Admoni, H., Matarić, M.: Robots for Use in Autism Research. Annu. Rev. Biomed. Eng. 14, 275–294 (2012).
13. Petric, F.: Robotic Autism Spectrum Disorder Diagnostic Protocol: Basis for Cognitive and Interactive Robotic Systems.
14. Petric, F., Miklic, D., Kovacic, Z.: Robot-assisted Autism Spectrum Disorder Diagnostics using POMDPs. Proc. Companion 2017 ACM/IEEE Int. Conf. Human-Robot Interact. - HRI '17. 369–370 (2017).
15. Wijayasinghe, I.B., Ranatunga, I., Balakrishnan, N., Bugnariu, N., Popa, D.O.: Human???Robot Gesture Analysis for Objective Assessment of Autism Spectrum Disorder. Int. J. Soc. Robot. 8, 695–707 (2016).
16. Kojima, H., Takaeda, K., Nihel, M., Sadohara, K., Ohnaka, S. & Inoue, T.: Acquisition and evaluation of a human-robot elderly spoken dialog corpus for developing computerized cognitive assessment systems. J. Acoust. Soc. Am. 140, 2963–2963 (2016).
17. Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. In: Journal of Internal Medicine. pp. 183–194 (2004).

18. Luis, C., Loewenstein, D., Acevedo, a, Barker, W.W., Duara, R.: Mild cognitive impairment: directions for future research. Neurology. 61, 438–444 (2003).
19. Landau, S.M., Harvey, D., Madison, C.M., Reiman, E.M., Foster, N.L., Aisen, P.S., Petersen, R.C., Shaw, L.M., Trojanowski, J.Q., Jack, C.R., Weiner, M.W., Jagust, W.J.: Comparing predictors of conversion and decline in mild cognitive impairment. Neurology. 75, 230–238 (2010).
20. Caraci, F., Castellano, S., Salomone, S., Drago, F., Bosco, P., Di Nuovo, S.: Searching for disease-modifying drugs in AD: can we combine neuropsychological tools with biological markers? CNS Neurol. Disord. Drug Targets. 13, 173–186 (2014).
21. Nasreddine, Z.S., Phillips, N.A., B??dirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H.: The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53, 695–699 (2005).
22. Chertkow, H., Nasreddine, Z.S., Johns, E., Phillips, N.A., McHenry, C.: The Montreal cognitive assessment (MoCA): Validation of alternate forms and new recommendations for education corrections. Alzheimer's Dement. 7, S157 (2011).
23. Xu, T. (Linger), Zhang, H., Yu, C.: See You See Me: The Role of Eye Contact in Multimodal Human-Robot Interaction. ACM Trans. Interact. Intell. Syst. 6, 1–22 (2016).
24. Sciutti, A., Rea, F., Sandini, G.: When you are young, (robot's) looks matter. Developmental changes in the desired properties of a robot friend. In: IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions. pp. 567–573 (2014).
25. Cohen, J.: Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum (1988).