



Administrating Cognitive Tests Through HRI: An Application of an Automatic Scoring System Through Visual Analysis

Sara Sangiovanni¹, Matteo Spezialetti², Fabio Aurelio D'Asaro¹(✉), Gianpaolo Maggi³, and Silvia Rossi¹

¹ Università degli Studi di Napoli Federico II, Naples, Italy
fdasaro@gmail.com

² Università degli Studi dell'Aquila, L'Aquila, Italy

³ Università degli Studi della Campania Luigi Vanvitelli, Caserta, Italy

Abstract. In this paper, we propose the use of a social robot to support professional figures in performing cognitive screening and stimulation. We implemented tools that allow us to automatize the evaluation of a subject's cognitive abilities. More specifically, we programmed the humanoid robot *Pepper* to administer, in a fully automated and “friendly” way, three cognitive assessment tasks: *Word List Recall* (WLR), *Attentive Matrices* (AM), and the *Rey-Osterrieth Complex Figure* (ROCF). For WLR, we displayed the word list on the robot's tablet, and the speech recognition module to record the recalled words. AM was delivered by asking the subjects to use and mark numbers on the tablet. For ROCF, we implemented two novel score assessment algorithms based on the processing of the picture drawn by the subjects and acquired through the robot camera. In particular, for ROCF, correlation analysis was conducted to compare automatically computed scores with a human psychologist's. Our results suggest that the human psychologist's workload can be reliably reduced thanks to the support of the robot.

Keywords: HRI · Cognitive tests · Automatic assessment

1 Introduction

A *social robot* is an autonomous robot capable of interacting with humans, that can follow social norms and role-specific rules. Currently, the use of social robots has become widespread in different areas, such as schools, transports, working contexts, and others. In this paper, we present the use of a social robot in a therapeutic setting, as support to professional figures for performing cognitive screening and stimulation. Research has found that cognitive stimulation can prevent the development of dementia [21]. Several studies explored the effectiveness of cognitive-based interventions and have shown that people undergoing cognitive training report a slower decline in daily activities (see e.g. [20]).

Currently, there are many technologies, including robotics ones, designed to submit cognitive training or to support professional figures in this area. Among these, social robots could represent a pleasant and understandable interface for facilitating the screening process. Social robots have been used to improve Autistic Spectrum Disorder (ASD) diagnosis in children [11] or to administer Patient-Reported Outcome measurement questionnaires to senior adults [3]. The success of this approach can be favored by the preference of interacting with a humanoid social robot rather than a non-embodied computer screen [15]. Moreover, results presented in [17] suggest that using social robots to support professional figures improves user performance, but the technology needs improvement for a fully autonomous assessment. Little is known about robots as psychological evaluation tools.

In our work, we explore the use of social robots to automate the psychological evaluation attributed to the performed cognitive exercises. Consequently, our main objective is to use a social robot to automatically deliver and evaluate cognitive tests. In particular, we implemented tools that enable a robot to automatically evaluate a subject's cognitive abilities. Among activities to be automatically performed by a robot, the computerized analysis of drawing-based test is still a complex task due to the high degree of drawing variability and the possible interpretation [7]. This activity is also time consuming for psychologists. Here, we propose two algorithms for the evaluation of the Rey–Osterrieth Complex Figure (ROCF) to be performed autonomously by a Pepper robot. Results show that is a strong correlation between our automatic evaluation methods and that of a human psychologist, but further analysis is required to fine tune our approach.

2 Related Work

The use of social robots in healthcare is a relatively new field of study. For instance, [10] explores the use of social robots in the therapeutic field. In particular, they are used in healthcare to provide monitoring, health education, and entertainment to patients.

Neurological examination provides important information on cognitive abilities of the therapeutic subject. Cognitive screening is the first step in neurological evaluation. Recent research like [18] demonstrates the benefits of using social robots as therapeutic assistants, as they can provide many advantages to diagnostic practice, for example by ensuring standardization in the subject's evaluation. Rossi et al. [14] explore the use of social robots as psychometric tools for providing quick and reliable screening exams. More in detail, this study compares the prototype of a robotic cognitive test to a traditional paper and pencil psychometric tool. Rossi et al. [15] have shown that subjects are likely to rate their experience as more satisfactory when they use a humanoid robot when compared to a mobile application. For example, the iCat robot recommends the use of ecological, energy-saving washing programs, while communicating through voice messages and facial expressions. The results suggest that people involved

in the experiment were more heavily influenced by the robotic cat rather than by the luminous information message on the washing machine display. Kidd and Breazeal [5] proposed a robotic trainer for weight loss for smart homes. Kidd and David in [6], where participants were asked to join a weight loss program, the authors found that the program was more effective when a robot was involved in the monitoring process when compared to both pen and paper and computerized interface-based approaches.

In this paper, we develop an algorithm for the evaluation of the *Rey–Osterrieth Complex Figure* (ROCF) that requires scoring a drawing in a human-like manner. Related work in this field uses fuzzy expert systems [4] that, however, does not provide a global evaluation of the figure due to localization issues. Other techniques are based on Deep Neural Networks [7]. A shortcoming of such methods is that they require a huge amount of data for the training phase. Our hybrid method, instead, provides a lightweight way to provide a global evaluation of the drawing using standard techniques from computer vision. As it is written in Python, our algorithm can be readily embedded in a Pepper robot and does not require any post-processing phase.

3 Materials and Methods

In our application, we used the humanoid robot *Pepper Y20 V18A*¹ and the *Choreograph suite*, which is included in Pepper’s SDK. Image processing operations were performed using Python and OpenCV libraries. We programmed Pepper to administer, in a fully automated way, the three cognitive assessment tasks listed below:

Word List Recall. *Word List Recall* (WLR) is used to evaluate learning and verbal memory abilities. In particular, the participants are shown 30 semantically unrelated words. After 1 min, the list disappears and the participants have to recall all the words they remember. The number of words correctly recalled is then calculated, and the participants are assigned a score from 0 to 30, with higher scores indicating better performance.

Attentive Matrices. To assess selective visual attention, we employed the *Attentive matrices* (AM) test [16]. It consists of three matrices with numbers arranged in a random sequence. The participant has to check the matrices to find and tick the target numbers shown at the top of the screen within 45 s (see Fig. 1). Then, the participant is assigned a score ranging from 0 (worst performance) to 60 (best performance).

Rey–Osterrieth Complex Figure. The *Rey–Osterrieth Complex Figure* (ROCF) is often used for the assessment of visuo-constructional and planning

¹ SoftBank Robotics <https://www.softbankrobotics.com>.



Fig. 1. Pepper performing the *Attentive Matrices Test* on its in-built tablet.

abilities, due to the complexity of the figure [19]. Participants are asked to copy the complex figure and show the sheet to Pepper (see Fig. 2). The image was then scored according to the number of elements present in the complex figure correctly copied so that higher scores indicate better performance (range 0–36). When the test is administered by a human psychologist, the 18 elements composing the ROCF are evaluated individually, applying a particular scoring grid.

3.1 Pepper's Behavioral Features

We programmed Pepper to behave in a friendly way. In the following, we describe the characterizing features of the friendly personality we adopted. Pepper's eye color was set to yellow. According to the color-wheel model by Plutchik yellow is typically associated with positive emotions such as joy and serenity [12]. Its gestures are frequent and open to give a sense of rhythm to its speech [8]. The openness of gestures is typically related also to the display of positive emotions [13]. Both the speed of speech and voice pitch are high, 90% and 100% of the default value respectively which are associated with a more entertaining robot [9]. For what regards the proxemics, we adopted Hall's personal space (0.3–1m) because reaches the right balance between greater persuasiveness and low discomfort: The language type used is informal to promote intimacy [2]. Finally, Pepper's gaze is fixed on the user during deliveries and distracted during the execution [2]. To enhance the perception of the robot's friendly personality, we adopted different motivational strategies: phrases of positive encouragement [1] are randomly repeated every 10 s, and a yellow smile is shown on Pepper's tablet during waiting times.

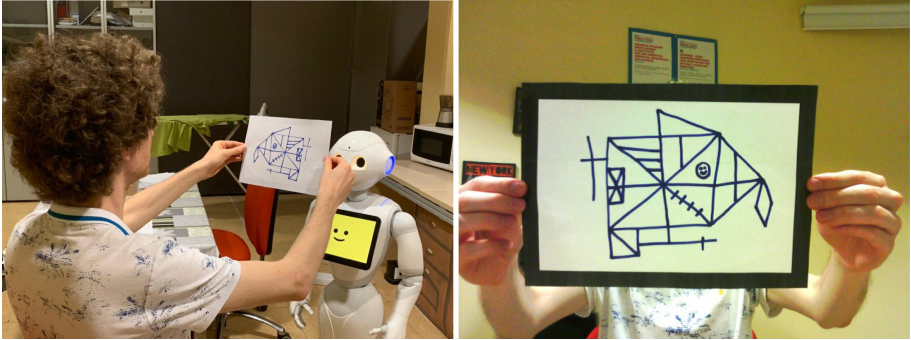


Fig. 2. A participant showing Pepper his reproduction of ROCF.

3.2 Automatic Cognitive Tests Administration

The interaction is entirely guided and supervised by Pepper, who first introduces itself and then explains the purpose of interaction to the participant. Subsequently, it proceeds to administer the cognitive tests, in the same way a human psychologist would, in the following order:

1. First phase: Pepper shows on the tablet the image of the original ROCF model, and asks the participant to copy the image on a white sheet. This task is meant to assess the subject's visuospatial and visuo-constructional abilities but also abilities of planning and organization. Pepper waits for the participant to say "I'm done" before continuing;
2. Second phase: in this phase, Pepper administers the two other proposed cognitive exercises (WL first, and then AM) in order to fill the interval before the ROCF delayed recall with other neuropsychological tasks;
3. Third phase: in this phase, the participant is asked to recall the ROCF. Particularly, through this task, it is possible to evaluate the subject's long term spatial memory. Pepper asks the participant to draw on a white sheet all the details s/he remembers of the original model shown in step one. At the end of the exercise, Pepper asks the participant to position the drawing in front of its eyes and takes a photo that is then used for automatic evaluation (see Fig. 2).

Finally, Pepper thanks the participant for the collaboration and ends the interaction. Note that all of our experiments were performed under controlled conditions. A human operator checked that the lighting conditions were fit for the task, and that Pepper's camera was correctly aligned to the sheet when capturing the image.

3.3 Automatic Cognitive Tests Evaluation

Word List Recall. To automatically administrate WLR, we used the tablet screen embedded in the robot to display the word list, and the speech recognition

module to record and take note of the recalled words. In case a word is recognized, Pepper repeats it for confirmation. The speech recognition module allows one to set a sensitivity value that represents how accurately the word has to be pronounced to be recognized. We set the sensitivity to 60% after a series of preliminary tests conducted to find the best trade-off between false positives (incorrect words recognized as correct, or correct words recognized as other words in the list) and false negatives (correct words not recognized). In particular, we looked for the lowest value that did not produce any false positive, since false negatives can be managed by simply ignoring the word.

Attentive Matrices. AM test was administered by showing the matrices on Pepper's tablet. The subject was asked to tick the numbers by tapping on them over the screen. The final score was the number of correctly ticked numbers.

Rey–Osterrieth Complex Figure. ROCF score assessment algorithm was based on the comparison between the image drawn by the subject and the one stored in the robot's memory to return a score. The sheet included a black frame (2.5 cm of thickness) on the border, ensuring that the subject did not cover the drawing area with fingers when showing the sheet to Pepper. To make the drawn image comparable with the original, the following preprocessing phase were applied:

- Binarization (see Fig. 3 for an example)
- Selection of the maximum bounding box of contiguous foreground pixels: each pixel outside the bounding box was set to 0
- Perspective linear transformation
- Border artifacts removal: each region of contiguous background pixel touching the border was set to foreground value 1.

The following steps were performed both on the preprocessed image and the original one. The aim was to compute an N by M similarity matrix that contained a score for each pair (O_i, D_j) of graphic elements (i.e. foreground connected regions) belonging to the original and the drawn image respectively:

- **Labeling:** contiguous pixels were labeled with the same integer value (see Fig. 3 for a pictorial representation). Regions were arranged in a list structure
- **Background removal:** the region with the larger bounding box area, was removed from the list
- **Removal of small regions:** regions composed of less than 50 pixels were not considered and removed from the list.

Both the shape similarity and the correct positioning of drawn elements with respect to the originals had to be quantified. Let O_i and D_j the i^{th} and j^{th} ($1 \leq i \leq N$, $1 \leq j \leq M$) elements of the regions lists from the original image and that drawn by the subject, respectively. For each pair (O_i, D_j) we compute the following similarity metrics:

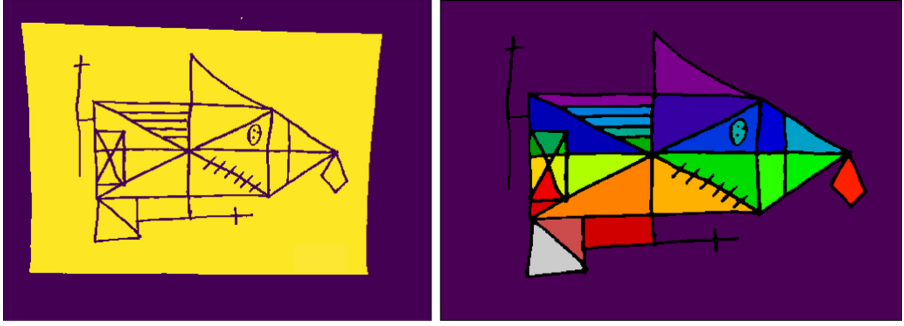


Fig. 3. A ROCF drawing after the binarization phase (left) and labeling phase (right). In the right picture, different colors correspond to differently labeled areas of the figure. (Color figure online)

Jaccard Index (JI): It is a value that takes into account the overlap of regions and it is defined as

$$JI(O_i, D_j) = \frac{S(O_i \cap D_j)}{S(O_i \cup D_j)} \tag{1}$$

where $S(x)$ represents the number of pixels in a region and set operators are applied by considering regions as sets of pixels;

Orientation Similarity (OS): Let be $\alpha_O, \alpha_D \in [0, \pi]$ the orientations of the major axes of the ellipses approximating the regions O_i and D_j , respectively. Let be R_O and R_D the ratios between lengths of major and minor axes of the ellipses. We computed the orientation similarity as follows:

$$OS(O_i, D_j) = \begin{cases} 0 & \text{if } (R_O \geq 0.5 \wedge R_D < 0.5) \vee \\ & (R_O < 0.5 \wedge R_D \geq 0.5) \\ 0.5 & \text{if } R_O \geq 0.5 \wedge R_D \geq 0.5 \\ (1 - |\sin(\alpha_O - \alpha_D)|) & \text{if } R_O < 0.5 \wedge R_D < 0.5 \end{cases} \tag{2}$$

Normalized Distance (ND): It is computed as

$$ND(O_i, D_j) = 1 - \frac{d(C(O_i), C(D_j))}{\max_{1 \leq n \leq N, 1 \leq m \leq M} d(C(O_n), C(D_m))} \tag{3}$$

where $C(X)$ is the coordinates vector of the centroid of region X and $d(y, z)$ is the Euclidean distance between points y and z .

Surfaces ratio (SR): It represents the similarity in terms of the number of pixels composing the regions:

$$SR(O_i, D_j) = \frac{\min(|O_i|, |D_j|)}{\max(|O_i|, |D_j|)} \tag{4}$$

where $|X|$ is the number of pixels belonging to region X .

The computed metrics were combined in a weighted sum and the N by M matrix A was filled as follows:

$$A(i, j) = \frac{w_1 JI(O_i, D_j) + w_2 OS(O_i, D_j) + w_3 ND(O_i, D_j) + w_4 SR(O_i, D_j)}{w_1 + w_2 + w_3 + w_4} \quad (5)$$

At this point, A was analyzed to compute a total score of the test. We developed two algorithms, named A1 and A2, respectively. Attention had to be paid in avoiding the “overloading” of drawn elements (when paired with several original regions). For these reasons, both algorithms implemented penalization mechanisms. Note that, since previous steps led to a division of the original image in 18 regions and each element of A was in the range $[0, 1]$, picking and summing an element of A for each original region, would produce a score in the range $[0, 18]$, thus the following algorithms also had to normalize the score to the range $[0, 36]$.

A1 comprised the following steps:

```
BEGIN
  score=0;
  WHILE max(A)>0
    find max(A) (let x and y be the indices);
    score = score + max(A);
    set each element in the x-th row of A to 0;
    multiply each element in the y-th column of A by PF;
  END
  RETURN 2*score;
END
```

At each step, it found the maximum score pair in the matrix and add the value to the final score. Then, to exclude the already considered regions between original, it set the corresponding row to 0 (in this way, each original region was paired with exactly one drawn region). Moreover, to avoid that a drawn region was selected repeatedly, the algorithm penalized it by multiplying the values in the corresponding column by a penalization factor $PF \in [0, 1]$. Setting $PF = 0$ is equivalent to make the algorithm behave with a hard constraint with respect to overloading: at most one original region will be paired with each drawn region, since, once a regions' pair is selected, the corresponding row and column is set to 0. On the contrary, $PF = 1$ means that no penalty will be applied for overloading. In this work, we tested the algorithm with values of $PF = 0, 0.5, 1$.

A2 is a simplification that addressed the correctness of the paired regions by using an overall metric, instead of computing penalty for every single region. The final score was computed as follows:

$$s = \sum_{i=1}^N (\max(A[i, :])) * P \quad (6)$$

where $P = \frac{2N}{N+(N-M)^2}$ was the weight term ($P \in [0, 2]$) that considered the number of original and drawn regions.

Table 1. Correlation scores (Pearson’s r , Spearman’s ρ , and Cronbach’s α) between our automatic evaluation algorithms of a task and the human psychologist’s evaluation. All results are significant ($p < 0.01$).

	r	ρ	α
A1 with $PF = 1$	0.73	0.82	0.71
A1 with $PF = 0$	0.79	0.84	0.87
A1 with $PF = 0.5$	0.75	0.86	0.85
A2	0.75	0.81	0.86

4 Evaluation and Results

We wanted to make sure our evaluation methods are reliable, in the sense that they must show a strong correlation with respect to a human psychologist’s evaluation. To this aim, we asked 37 participants (19 male, 18 female, aged 23–38) to take the three tests (WL, AM, and ROCF) under the supervision of a human. These tests were evaluated by both a psychologist and by our algorithms. We obtained 100% accuracy when scoring WL and AM, as these tests are performed using the robot’s tablet and speech recognition software. Thus, in the following, we focus on ROCF. We tested algorithm A1 for PF equal to 0, 0.5 and 1, and algorithm A2. Then, we compared the expert’s scores to the automatically calculated ones.

Table 1 shows two measures of correlation between our algorithms and the expert’s evaluation. It is worth noting that A1 with PF equal to 0 and 0.5 show higher correlation scores when compared to A2.

To further investigate our algorithms for ROCF evaluation we also did regression analysis using a linear model (see Fig. 4). The slope test for all models rejected the hypothesis that they have a slope equal to 0 ($p < 0.001$), again implying correlation with the human psychologist’s evaluation. Unfortunately, the slope test also rejects the hypothesis that models of A1 with PF equal to 0, 0.5, and 1 have a slope equal to 1 ($p < 0.01$). This hypothesis is *not* rejected for A2, which possibly indicates that A2 is the best choice as an automatic evaluation metric for ROCF straight out-of-the-box. On the other hand, A1 with $PF = 0$ and $PF = 0.5$ have higher correlation scores than A2. This seems to suggest that these two methods only differ from the expert’s evaluation by a scale/rotation factor and, if appropriately tweaked as to systematically adjust their slope, might overperform A2. More data is needed to verify this hypothesis and use it to improve our method. We intend to do so in future work.

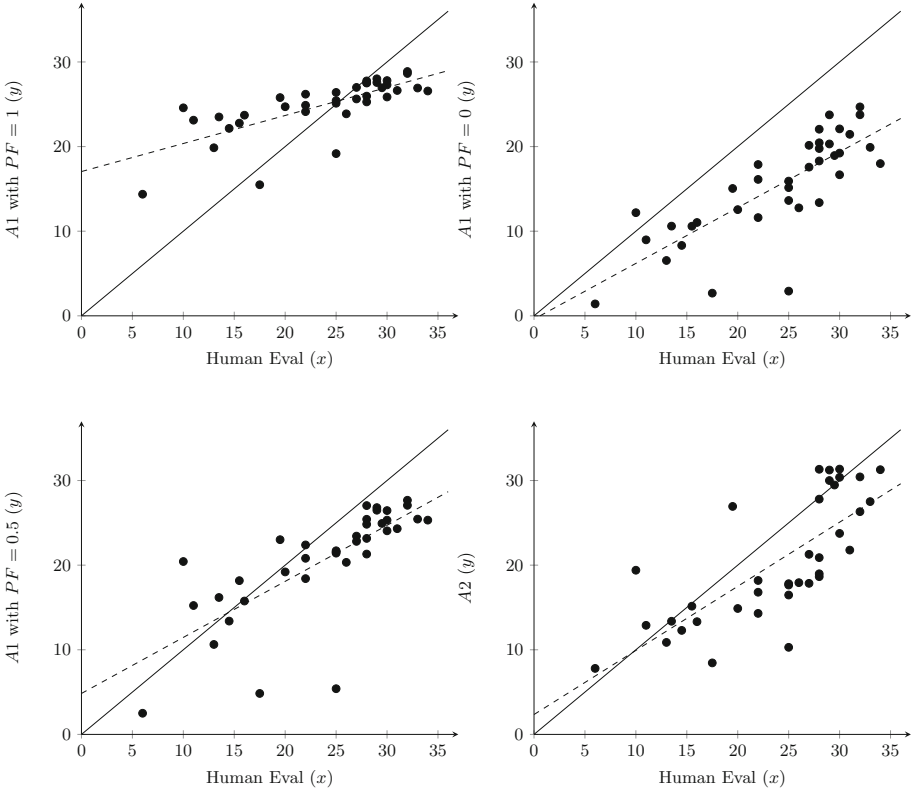


Fig. 4. Linear models of our two algorithms. The equations of the four models (dashed lines) are: $y = 0.33x + 17.05$ (A1 with $PF = 1$, shown in Fig. 4a) with $R^2 = 0.54$, $y = 0.66x - 0.39$ (A1 with $PF = 0$, shown in Fig. 4b) with $R^2 = 0.62$, $y = 0.66x + 4.84$ (A1 with $PF = 0.5$, shown in Fig. 4c) with $R^2 = 0.56$, $y = 0.76x + 2.35$ (A2, shown in Fig. 4d) with $R^2 = 0.56$. For reference, we display a $y = x$ solid line in every plot.

5 Conclusion and Future Work

In this work, we developed and implemented a series of tools that aim to automatize the administration and evaluation of cognitive tests. These tasks are typically carried out by a human psychologist and a psychologist. However, we demonstrated that the human operator can be reliably supported by an HRI system (a Pepper Robot, in our application). Indeed, our results suggest that there is a strong correlation between our automatic evaluation methods and that of a human psychologist. In particular, our implementation of algorithms to automatically score the Rey-Osterrieth Complex Figure showed a high correlation with scores assigned by a psychologist. In future work, we plan to improve these algorithms by collecting more data, which would allow us to fine-tune the automatic evaluations, and refine the recognition of finer details.

It is worth noting here that participants in our pilot study are healthy young adults (aged 23–38). This constitutes a limitation of our work, as older/cognitively impaired individuals may reproduce highly distorted ROCF figures that may be harder for our algorithms to correctly evaluate. As previously noted, another source of image distortion may be due to user-independent factors such as non-uniform lighting conditions, partially acquired images, misaligned camera, etc. We did not observe such limit conditions here, as a human operator always made sure the environmental conditions were fit. We plan to further investigate these possible limitations in future work.

Another objective for future work is to contribute an open dataset for the evaluation of the ROCF. We believe a standard set of figures captured from Pepper’s camera would greatly help comparing different algorithms for the evaluation of ROCF, and would be beneficial for Social Assistive Robotics studies that aim to automatize similar tasks.

These findings may prove useful for further development of similar fully autonomous agents for the administration of tests, to be employed e.g. in health-care services and interactive cognitive training. Our tools may also turn useful to *standardize* the evaluation of tests, as they are independent of external factors that may affect a human operator, such as personality and subjective marking criteria.

Acknowledgments. This study was partially funded by MIUR (Italian Ministry of Education, Universities, and Research) within the PRIN2015 research project UPA4SAR - User-centered Profiling and Adaptation for Socially Assistive Robotics (grant n. 2015KBL78T).

References

1. Agrigoroaie, R., Tapus, A.: Influence of robot’s interaction style on performance in a Stroop task. In: Kheddar, A., et al. (eds.) *Social Robotics*. LNCS, vol. 10652, pp. 95–104. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-70022-9_10
2. Baxter, P., Ashurst, E., Read, R., Kennedy, J., Belpaeme, T.: Robot education peers in a situated primary school study: personalisation promotes child learning. *PLoS ONE* **12**(5), e0178126 (2017)
3. Boumans, R., Van Meulen, F., Hindriks, K., Neerinx, M., Olde Rikkert, M.: Proof of concept of a social robot for patient reported outcome measurements in elderly persons. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 73–74 (2018)
4. Canham, R., Smith, S.L., Tyrrell, A.M.: Automated scoring of a neuropsychological test: the Rey Osterrieth complex figure. In: *Proceedings of the 26th Euromicro Conference. EUROMICRO 2000. Informatics: Inventing the Future*, vol. 2, pp. 406–413. IEEE (2000)
5. Kidd, C.D., Breazeal, C.: A robotic weight loss coach. In: *Proceedings of the National Conference on Artificial Intelligence*, vol. 22, p. 1985. AAAI Press/MIT Press. Cambridge/Menlo Park (2007)
6. Kidd, C.D.: *Designing for long-term human-robot interaction and application to weight loss* (2008)

7. Moetesum, M., Siddiqi, I., Ehsan, S., Vincent, N.: Deformation modeling and classification using deep convolutional neural networks for computerized analysis of neuropsychological drawings. *Neural Comput. Appl.* **32**, 1–25 (2020). <https://doi.org/10.1007/s00521-020-04735-8>
8. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) *IVA 2010. LNCS (LNAI)*, vol. 6356, pp. 222–235. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15892-6_24
9. Niculescu, A., van Dijk, B., Nijholt, A., Li, H., See, S.L.: Making social robots more attractive: the effects of voice pitch, humor and empathy. *Int. J. Soc. Robot.* **5**(2), 171–191 (2013). <https://doi.org/10.1007/s12369-012-0171-x>
10. Olaronke, I., Oluwaseun, O., Rhoda, I.: State of the art: a study of human-robot interaction in healthcare. *Int. J. Inf. Eng. Electron. Bus.* **9**, 43–55 (2017)
11. Petric, F., Kovacić, Z.: No data? No problem! expert system approach to designing a POMDP framework for robot-assisted ASD diagnostics. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 209–210 (2018)
12. Plutchik, R.: The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am. Sci.* **89**(4), 344–350 (2001)
13. Rossi, S., Dell’Aquila, E., Bucci, B.: Evaluating the emotional valence of affective sounds for child-robot interaction. In: Salichs, M.A., et al. (eds.) *ICSR 2019. LNCS (LNAI)*, vol. 11876, pp. 505–514. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35888-4_47
14. Rossi, S., Santangelo, G., Staffa, M., Varrasi, S., Conti, D., Di Nuovo, A.: Psychometric evaluation supported by a social robot: personality factors and technology acceptance. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 802–807. IEEE (2018)
15. Rossi, S., Staffa, M., Tamburro, A.: Socially assistive robot for providing recommendations: comparing a humanoid robot with a mobile application. *Int. J. Soc. Robot.* **10**(2), 265–278 (2018). <https://doi.org/10.1007/s12369-018-0469-4>
16. Spinnler, H.: Standardizzazione e taratura italiana di test neuropsicologici. *Ital. J. Neurol. Sci.* **6**, 21–120 (1987)
17. Varrasi, S., Di Nuovo, S., Conti, D., Di Nuovo, A.: A social robot for cognitive assessment. In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 269–270 (2018)
18. Varrasi, S., Di Nuovo, S., Conti, D., Di Nuovo, A.: Social robots as psychometric tools for cognitive assessment: a pilot test. In: Ficuciello, F., Ruggiero, F., Finzi, A. (eds.) *Human Friendly Robotics. Springer Proceedings in Advanced Robotics*, vol. 7, pp. 99–112. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-319-89327-3_8
19. Watanabe, K., et al.: The Rey-Osterrieth complex figure as a measure of executive function in childhood. *Brain Dev.* **27**(8), 564–569 (2005)
20. Willis, S.L., et al.: Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA* **296**(23), 2805–2814 (2006)
21. Wilson, R.S., et al.: Participation in cognitively stimulating activities and risk of incident Alzheimer disease. *JAMA* **287**(6), 742–748 (2002)