Assessment by telematic means and artificial agents: A new challenge for psychometrics?

Daniela Conti[1,2], Santo Di Nuovo[3]

[1] *Department of Humanities, University of Catania, Catania, Italy*
[2] *Oasi Research Institute-IRCCS, Troina (EN), Italy*
[3] *Department of Educational Sciences, University of Catania, Catania, Italy*

daniela.conti@unict.it

♟ ABSTRACT. L'utilizzo di strumenti psicodiagnostici somministrati elettronicamente e/o con agenti artificiali ha aperto una nuova sfida. A un campione di 122 psicologi abilitati è stato somministrato uno strumento che comprendeva un questionario composto da 21 items su una scala a 5 punti e un Differenziale semantico. L'atteggiamento generale degli psicologi nei confronti delle modalità innovative di valutazione è risultato positivo. Dall'analisi dei dati emerge come la somministrazione di test psicodiagnostici per via elettronica o con agenti artificiali richiede studi di adattamento e in molti casi una riformulazione degli strumenti e una adeguata formazione dei professionisti.

♟ SUMMARY. *During the COVID-19 restrictions, the administration of psychodiagnostic tools not through direct interaction with the psychologist, but administered electronically and/or with artificial agents, opened a new challenge. A change of setting and interactive conditions are introduced which can alter the reliability and validity of tools consolidated for diagnostic use in face-to-face assessments. A sample of 122 licensed psychologists participated in the study. They were either attending or teaching post-graduate specialization courses of different theoretical-methodological focus. The participants were given an online survey via Google forms that included a questionnaire composed of 21 items on a 5-point scale, built according to the UTAUT model and adapted for the acceptability and willingness to use online questionnaires for mental well-being, and a semantic differential for evaluating the attitude towards technology in general. The overall attitude of the psychologists toward the innovative modalities of assessment is positive, and the intention to use online testing is very high. The predictors of the positive attitude and intention of use are analyzed. A multidimensional analysis suggested that the attitude towards online testing, and the intention to use it, are located in the crossing dimensions of technical (psychometric) aspects and concrete usability. The administration of tests electronically or through artificial agents requires adaptation studies and in many cases a reformulation of the tools that are offered using these modalities. The need to raise awareness amongst psychologists about the new forms of assessment, and to train those who intend to use them, has be underlined.*

Keywords: *Online assessment, Psychometry, Diagnostic process, Artificial agents*

INTRODUCTION

The recent pandemic has prompted the generalization of a practice that was already spreading, albeit amidst many doubts and contrasts: the administration of psychodiagnostic tools not through direct interaction with the psychologist, but administered electronically and/or with artificial agents.

This methodological scenario is very different from the use of technologies in the administration and evaluation of psychological tests.

The well-known *Computerized Adaptive Testing* allows tailoring of the instruments, and automating their adaptation to the subject's skill starting level, thus making tests more appropriate and precise, especially for extreme cases (Wainer, Dorans, Flaugher, Green & Mislevy, 2000).

Artificial intelligence systems facilitate complex diagnoses. As a recent example, Grazioli and colleagues (2022) developed supervised machine-learning algorithms to support the diagnostic process for attention deficit hyperactivity disorder (ADHD) in children. A decision tree classifier and random forest algorithms were used to identify the most relevant features in discriminating between the presence or absence of the ADHD diagnosis; they reliably aid the diagnostic process as an initial step.

However, in these cases the psychologist still administers the tests, while the conditions of adaptation to the user and the methods of scoring or interpretating the results have changed, now being completely automated.

When providing for remote administration, a change of setting and interactive conditions are introduced which can alter the reliability and validity of tools consolidated for diagnostic use in face-to-face assessments. Online versions of tests may not always measure the same constructs as their traditional antecedents, therefore equivalence cannot be assumed (Buchanan, 2002). The psychologists themselves may not adequately realize this difference.

Even the use of artificial agents (such as humanoid robots) for administering tests, which has been growing in recent years thanks to the advances in artificial intelligence systems based on Recursive Neural Networks, poses technical problems. These concern the psychometric properties of reliability and validity, and also problems of usability and acceptability of the technological modalities by both professionals and users. In fact, even when the relationship remains remote, and is not limited to the administration of the questionnaire online or in asynchronous mode, the remote modality cannot fully replace the interactive aspects, both verbal and non-verbal, which are relevant for the diagnosis in face-to-face situations.

We will summarize the results of research aimed at deepening the psychological and psychometric aspects of assessment using artificial agents (i.e., humanoid robots), telematic devices, or both.

The robot as a tool for psychometric diagnosis

Recent literature has shown that robots have the potential to be successful assistants in psychological assessment (Conti, Commodari & Buono, 2017; Di Nuovo et al., 2019; Fiske, Henningsen & Buyx, 2019; Riches et al., 2022; Rossi et al., 2018, 2020; Varrasi, Di Nuovo, Conti & Di Nuovo, 2018a, 2018b). However, when considering this method of administration using robots, perhaps the most innovative findings recently presented in the international literature on testing, show that there are relevant empirical results, even if still preliminary.

The literature on robotics that tries to address psychometric evaluation is quite limited. This is probably due to the skepticism of the practitioners (Conti, Di Nuovo, Buono & Di Nuovo, 2017) and to the perception of users on the reliability shown by the robots (de Graaf, Ben Allouch & van Dijk, 2019). Specifically, robots can be programmed to perform specific, repeatable actions, providing the benefit of attainable standardization. Therefore, the robotic implementation of quick screening tests could be promising, because they are often repetitive and easy to take, but can be time-consuming for human assessors (Di Nuovo et al., 2019). An example where this would work is the observation of developmental history and social skills, where clinicians with different specialisations often do not agree when evaluating the same patients (Scassellati, Admoni & Matarić, 2012).

However, since an essential characteristic of psycho-diagnostic tests is that the stimuli and the methods for their administration should be just as rigorously standardized to guarantee the reliability (i.e. its repeatability in different times and places) and the validity of the results, assistive robots can represent a valuable way to meet these requirements and provide a reliable automatic tool for psychometric assessment (Di Nuovo et al., 2019; Rossi et al., 2018; Varrasi et al., 2018b; Varrasi, Lucas, Soranzo, McNamara & Di Nuovo, 2019). A robot-led assessment can provide a series of advantages,

including: assessor neutrality, objective measurement of social behaviour, standardization of the interaction, and better acceptance of the robotic platform than a non-embodied computer (Feingold Polak et al., 2018; Varrasi et al., 2018a).

In a recent study (Di Nuovo et al., 2019) aimed at the preliminary evaluation of the feasibility of entire or partial robotic testing, the neuropsychological test *Montreal Cognitive Assessment* (*MoCA*) was administered to a pilot sample of 16 adults (age range 19-61) using the Pepper robot. The purpose was to compare the standard score obtained traditionally, using paper-pencil assessments, with the automatic score obtained by the software integrated with the robot, and the supervised score, calculated by a psychologist correcting the automatic score through audio-video analysis (thus evaluating the communication and relationship aspects). In this case, the score was more correlated with the standard version, demonstrating to the psychologist the need to supervise automated assessments. The analyzes were repeated for the individual subtests of the MoCA, demonstrating that the different tasks within the test are affected differently by the method of robotic administration.

Another study conducted by Rossi et al. (2020) on a sample of 19 adults (8 males, 11 females, age range 53-82) involved both the traditional paper-pencil assessments of the MoCA test and robotic assessments, in addition to the of NEO-PI-3 personalities and the UTAUT questionnaire to evaluate the acceptance of computerized procedures. The results showed that the personality trait *Openness to experience* influenced the performance obtained during the interaction with the robot, and facilitated the unsupervised application of the robotic tool for cognitive assessment. Furthermore, anxiety or empathy towards robotic assessment also influence the differences from traditional assessment.

For these reasons, an important objective of future research will be to evaluate the reliability and validity of the use of the humanoid robot as a tester by acting remotely on its control. The overall aim will be to use more robots in different contexts with centralized supervision and therefore a considerable expansion of the testing possibilities for large screenings and patients with frailty that need extra support or care.

Moreover, the possibility of using a virtual robot for testing at distance should be considered.

Baig and Kavakli (2019), presenting a review of methods and measurements that are currently used in psychophysiological analysis to measure cognitive or mental states, also report on techniques of virtual reality and telerobotics, concluding that virtual reality simulations can be used to study the relationship between brain responses and stress levels. A telepresence robot was designed to measure the physical and psychological health of elderly people (Cortellessa et al., 2018). The results of recent reviews (Virkus, Leoste, Marmor, Kasuk & Talisainen, 2023; Zhang & Hansen, 2022) suggest that from a psychological perspective, telepresence robots could enable greater social presence in remote communication, but that their acceptance for wider use is still challenged.

Studies on online assessment with telematic tools

Currently there are few empirical types of research aimed at evaluating assessment using telematic tools, despite an increase in tools that are not limited to automatic scoring, but also provide for the possibility of remote administration.

Since the 1980s, e-marking has allowed administering tests to large groups at the same time in the fields of educational and industrial training within e-learning paradigms (e.g., Bukie, 2014; Ivanović & Jain, 2013; Preston & Shackelford, 1999). The diffusion of telematic testing has led to the critical discussion of technical aspects: e.g., copyright of the material, modification of administration procedures, the need for dedicated platforms, *ad hoc* standards and thresholds that compensate for the measurement error components introduced by the telematic medium. Furthermore, by reducing the interaction between the psychologist and the user, the relational and non-verbal aspects which are useful for the global understanding of the examined subject are underestimated. In this way, essential components of the diagnostic work would be taken away from the psychologist, especially when working with patients with frailty that need extra support or care and children, or in diagnoses of forensic relevance, which could hardly be adapted to the different remote settings.

Some experimental research has studied the equivalence between standard and remote administration, for example by comparing the traditional and telematic versions of the *General Health Questionnaire – 28* and of the *Symptoms Check-List – 90 – R* (Vallejo, Jordán, Díaz, Comeche & Ortega, 2007).

A meta-analytic review (Brearly et al., 2017) compared neuropsychological tools administered face-to-face and telematically, without highlighting significant differences for tests that use the verbal channel, such as the *Mini-Mental State Examination* (*MMSE*), the digit span, the Boston Naming Test and the fluency test. Few experimental verifications exist regarding tests where the manipulation of complex perceptual stimuli is necessary.

Furthermore, the opportunity to use remote neuropsychological assessment tools for elderly patients or patients with mobility difficulties has been studied (Cullum, Hynan, Grosch, Parikh & Weiner, 2014; Wadsworth et al., 2018). The concrete problem is choosing tests that are more or less suitable for remote administration, using those whose basic constructs are best designated to the type of user and the objectives of the evaluation. The elderly, children and adolescents, family systems or couples all have very different characteristics for which telematics testing needs to be specifically adapted.

As with all technologies applied to psychology, remote assessment is not a solution for all logistical or organizational problems. It poses the problems of usability and acceptability that should demonstrate which techniques are most appropriate to be translated profitably into the remote format, acceptable to both users and psychologists.

On these issues, previous survey research (Di Nuovo & Narzisi, 2021) studied the acceptance by potential users of telematic interventions (including assessment) by a comparison group of psychologists. The results show that a greater acceptance of being evaluated and treated online is linked - in addition to the possession of adequate tools and technical skills - to the belief in obtaining health benefits. The two personality traits (evaluated using the Big Five Factors model) that most predispose potential users to the use of technology and telematics are Extroversion and Open-mindedness. These results are in line with the international literature on the subject.

However, some important concerns are raised by the answers from professional psychologists. Their positive evaluation of online assessment correlates with the perception of being able to use it with children and adolescents too, with the understanding that tests to be administered remotely require specific standardization and calibration. However, a more positive assessment of psychologists is also associated with a greater belief that specific preparation is not necessary to administer online tests. It seems that the technical

efficiency of the tool (either the telematic questionnaire or the robot tester) may be enough to ensure its diagnostic effectiveness.

Aims of the study

Reassuming the results of the literature, the need emerges for a further study aimed at deepening not only the general acceptance of the possibility of testing that does not foresee the direct relationship with the psychologist, but, more specifically, the technical conditions deemed necessary for the acceptance of these online testing methods by the professional psychologist.

Based on the preliminary purpose of the study, an exploratory approach was used. The specific aims are:

1. To verify the degree of acceptance of different aspects of telematic assessment, and the intention to use it, comparing these variables in the subgroups of psychologists based on gender and age.
2. To correlate the acceptance of remote assessment with the frequency of previous use of online questionnaires and with positive evaluation of the technology.
3. To ascertain what aspects of acceptance are more predictive of the intention of use.
4. To find dimensional relationships between the more relevant variables emerging in the study.

MATERIALS AND METHODS

Instruments

– The first instrument was a questionnaire involving the collection of responses to 21 items on a 5-point scale (see Appendix). Some of these items had been built according to the UTAUT model (Unified Theory of Acceptance and Use of Technology; Venkatesh, Morris, Davis & Davis, 2003), and had been adapted for the acceptability and use of online questionnaires for mental well-being (e.g., Sora, Nieto, Del Campo & Armayones, 2021). A specific section covered online psychodiagnostic testing. Here, 13 items were identified from a survey of online psychological assessments, promoted during the pandemic by the European Federation of Psychologists Associations.
Three additional areas of enquiry looked at socio-

demographic information (gender and age) and the previous use of online instruments (never used; rarely; sometimes; often used).

Preliminarily the questionnaire was verified using a panel of judges, i.e. four psychologists who were experts in psychometrics, who evaluated the comprehensibility of each item and its pertinence for the study. Afterwards, the critical items were reformulated, and only items with an interrater agreement for suitability ≥.80 were maintained in the final version. A specific question was dedicated to the intention to use online testing in own professional practice, to be rated on a 5-point scale, from *not at all* to *surely*.

- Moreover, a s*emantic differential* (SD) was administered, consisting of 12 pairs of polar adjectives (e.g.: easy/difficult; useful/useless) on which to evaluate technology in general (Technology is…). For each pair, 6 intervals of response were foreseen, 3 for each polarity; e.g. for easy/difficult: very easy, quite easy, not very easy, not very difficult, quite difficult, very difficult. This tool allowed for a more connotative than denotative evaluation of the research object, and therefore was more linked to emotionality than to purely cognitive judgment (Osgood, Suci & Tannenbaum, 1957; Takahashi, Ban & Asada, 2016). A preliminary analysis of the instrument in a previous study (Di Nuovo & Narzisi, 2021) has demonstrated its monofactoriality, based mainly on the evaluation factor, and the consequent possibility of obtaining a total score by adding the scores of the individual scales, after having rotated all the items in a positive direction. The analysis of the reliability of the total score showed sufficient internal consistency: Cronbach's alpha index was .68 in the pilot study.

Participants

A sample of 122 licensed psychologists participated in the study. They were either attending or teaching post-graduate specialization courses of different theoretical-methodological focus (i.e., Cognitive-behavioral, Gestalt and Psychodynamic).

The gender of the sample (M = 40: 32.79%; F = 82: 67.21%) reflects the different proportions existing in the population of psychological professionals. In terms of age groups, 52.46% (*N* = 64; M = 19, F = 45) were under 45; while 47.54% (*N* = 58; M = 21, F = 37) of the sample was aged 46 or older.

Participants came from various Italian regions: Lombardy, Veneto, Liguria, Lazio, Campania, Apulia, Sardinia, and Sicily.

Regarding the previous use of online tests and questionnaires, 40.98% declared that they had never used them; while 15.57% used them on rare occasions, 32.79% occasionally, and 10.66% frequently.

Procedure

The participants were given an online survey via Google forms that included the questions shown in Appendix (Italian version was administered). The sample was recruited from lists of Italian psychologists from different psychotherapy schools and their teachers/supervisors.

The participants were invited to take part in the research via email and subsequently gave online written informed consent to the aggregated use of their data and answered the questionnaire anonymously. They were informed that the questionnaire would take ~10-15 minutes to complete, and the instructions indicated that they could withdraw at any time.

The study was approved by the Ethics Committee of the Department of Educational Sciences of the University of Catania (Italy).

RESULTS AND DISCUSSION

Descripive statistics and subgroup comparison

The first analysis calculated the level of agreement for each item on the questionnaire across the total sample (see Figure 1). Figure 1, also shows the differences between gender and age, where significant on the *t*-test ($p < .05$).

In the psychologists' sample the items most agreed upon concerned the possession of adequate technological resources for online testing, along with the contribution of the pandemic to the spread of this practice, and the possibility of taking the test at times and in places most convenient to the customer (predominant in male respondents). These are attitudes concerned with usability, and were followed by concerns about technical conditions (problems with complex perceptive materials, difficulty to administer remotely, or with clients with frailty or children). The sample interviewed

D. Conti, S. Di Nuovo

Figure 1 – Rank of means of the items of the questionnaire in the total sample
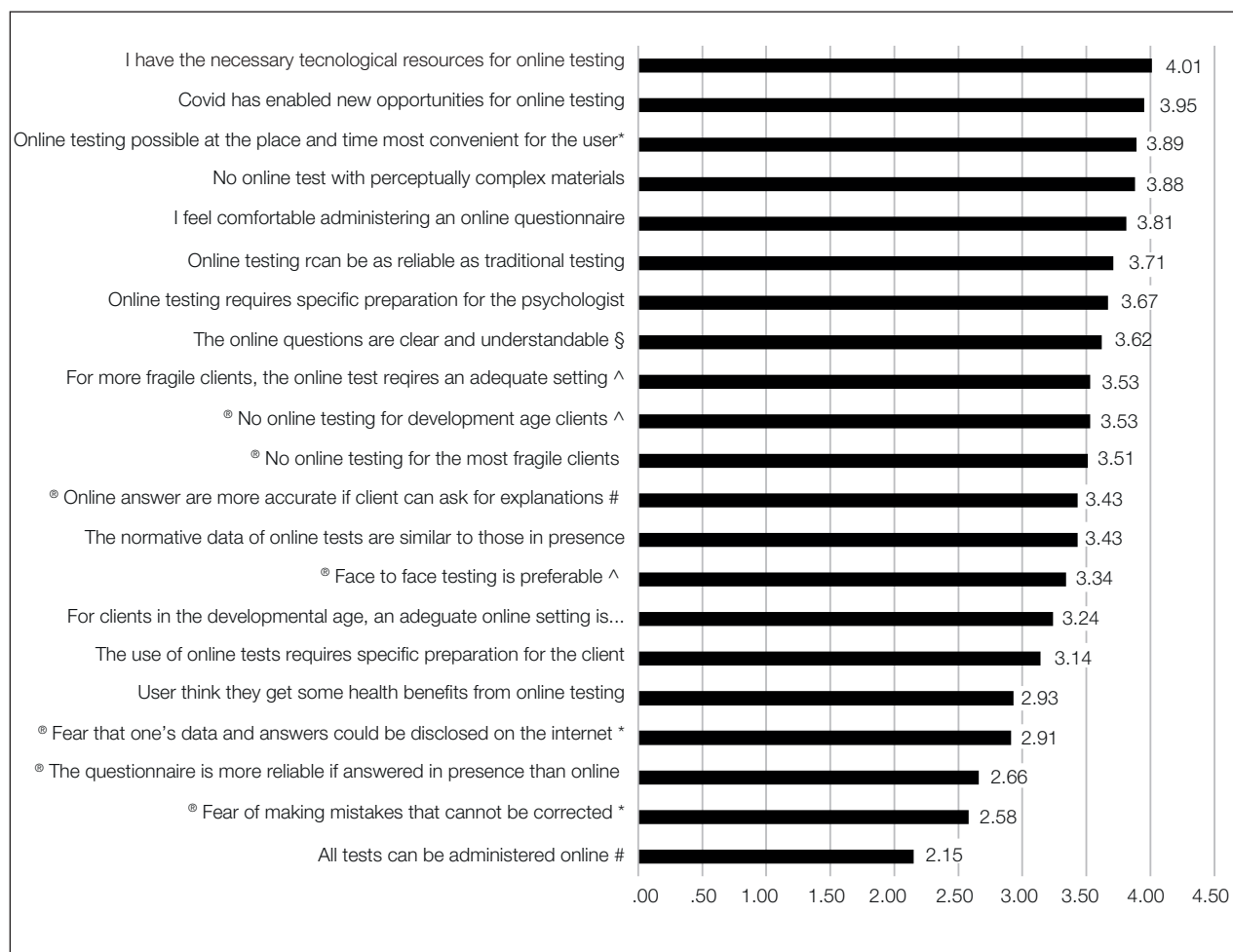
Note. Significant differences for gender and age are reported.

Legenda. ® = Reversed item; significant differences for gender: * = M>F, ^ = F>M; significant differences for age: # = older>younger, § = younger>older.

seems to be quite aware of these technical problems. However, there is a high level of agreement on the possibility that online testing can be as reliable as traditional tests, whilst recognising the need for specific preparation, not only for the patients, but also for the psychologists who use this method. However, 25.40% ($N = 31$) of the sample - regardless of gender and age - disagreed with the need of a specific formation.

The scoring for face-to-face testing being preferable, is quite high: the mean is 3.34 on the 5-point scale, and 45.08% of psychologists ($N = 55$) would prefer to administer the test in person, therefore considering online testing as a fallback. The

positive attitude is more present among older psychologists, where the percentage of agreement is 48.28%, against 42.19% of the younger ones.

Little importance is attributed to concerns about the possible violation of privacy, the difficulty of correcting any errors, or technical problems that are not considered more serious than in face-to-face testing.

As expected, the statement that "All tests can be used online" garnered the lowest ratings on the scale of agreement. The average score is very low (2.15 on a 5-point scale), but looking more closely at the single answers, one third of the

total (33.33%) are psychologists who agree (only 15, i.e. 12.50%, "completely agree"). The agreement is predominant at the higher age level ($\chi^2 = 12.61$, $df = 3$, $p = .01$).

On the other hand, the limited attribution to customers who trust being able to receive health benefits from online testing is surprising; only 26.23% of psychologists ($N = 32$) agree with this attribution, while the majority are in doubt or even against it, showing a certain distrust on the users' appreciation of the potential benefits (which instead resulted high in the research on users: see Di Nuovo & Narzisi, 2021).

A noticeable variable in our study is the intention to use online testing, which is quite high, but manifests a significant difference by age: youngest group $N = 64$, *Mean* = 4.44, *SD* = .81; older group $N = 58$, *Mean* = 3.88, *SD* = .97; $t_{120} = 3.44$, $p<.001$.

More than half of the younger respondents (59.38%) fully agree with the use of online testing, while only 10.94% refuse or are doubtful. In the group of older psychologists, the percentages are respectively 31.03 of full agreement and 32.76 of doubt or disagreement (the overall χ^2 is 12.74, $df = 3$, $p = .01$). But none of the respondents is completely against using the tests online.

No difference was found regarding gender ($\chi^2 = 1.08$, $df = 3$, $p = .78$).

The overall positivity towards technology assessment (total score derived from the semantic differential) is also, as expected, higher in the younger psychologists: younger group $N = 64$, *Mean* = 54.97, *SD* = 4.47; older group $N = 58$, *Mean* = 52.57, *SD* = 6.20; $t_{120} = 2.47$, $p<.01$.

Correlational analyses

The zero-order correlations between questionnaire items and frequency of previous use of online questionnaires and positive assessment of the technology were calculated using Pearson's coefficient. We report below on the correlations found to be significant for at least $p<.05$, after Bonferroni's correction for multiple comparisons.

The agreement with the statement: "Online testing is possible at the place and time most convenient for the user" correlates both with the frequency of previous use ($r = .40$, $p<.01$) and with the positive evaluation of the technology ($r = .24$, $p<.05$). The positive attitude towards technology correlates also with the agreement on the comprehensibility of the questions asked online ($r = .34$, $p<.01$), the recognition

of the input given by Covid to online testing ($r = .24$, $p<.05$), and the need for an adequate setting for the online testing of children ($r = .20$, $p<.05$).

As shown in Table 1, the predictivity of the questionnaire items for the willingness to use online tests was evaluated by a backward stepwise regression analysis.

The most predictive variables of the intention to use, i.e. those selected for the stepwise regression, are the absence of a prejudicial preference for face-to-face testing, the recognition of the usefulness deriving from the pandemic period, and the admission that the online test can also be useful for children (but with attention to specific settings for these age levels). The reduced concern about the difficulty of administering online perceptually complex stimuli is also relevant, along with the risks of privacy violations. Finally, the intention to use online testing is influenced by feeling comfortable in this way of working and the convenience of using the test at the most appropriate times and places for the client.

Multidimensional analysis

To find dimensional relationships between the more relevant variables emerging in the study, the correlation matrix among the variables significant at the stepwise regression analysis, together with the three core variables (Intention of use, Previous use, Positive attitude toward technology), was submitted to of multidimensional scaling in two dimensions, using the Guttman/Lingoes method. Figure 2 shows the graphical results of the analysis.

Dimension 1 contrasts the attitudes towards online testing, showing the positive values on the right side of the graphic with the negative aspects on the left side. The positive aspects (i.e. experience during the pandemic, the possibility of specific settings, and feeling comfortable with online testing) are linked to the positive attitude towards technology and to the intention of use.

Dimension 2 contrasts technical aspects, both positive and negative (upper part of the plot), to the aspects related to the use (lower half of the graphic).

The analysis clearly shows that the technical aspects and usability are intertwined in the perception by the psychologists for defining the dimensions of online testing and the actual intention of use. Previous use of online instruments is neutral for both dimensions, although closely linked to the availability for use.

Table 1 – Backward stepwise regression

	St.coeff.	*p*
® Testing preferable in presence	−.38	<.001
Covid practice as input	.33	<.001
Specific setting needed for developmental age	.20	.04
® No for perceptual stimuli	−.19	.05
® Fear for data diffusion	−.34	.07
® No for developmental age	−.20	.09
Comfortable with online testing	.14	.10
Facility of place/time for users	.12	.10

Note. Dependent variable: intention of use, *p*-to-remove = .15, (R^2 = .47).

Figure 2 – Multidimensional scaling
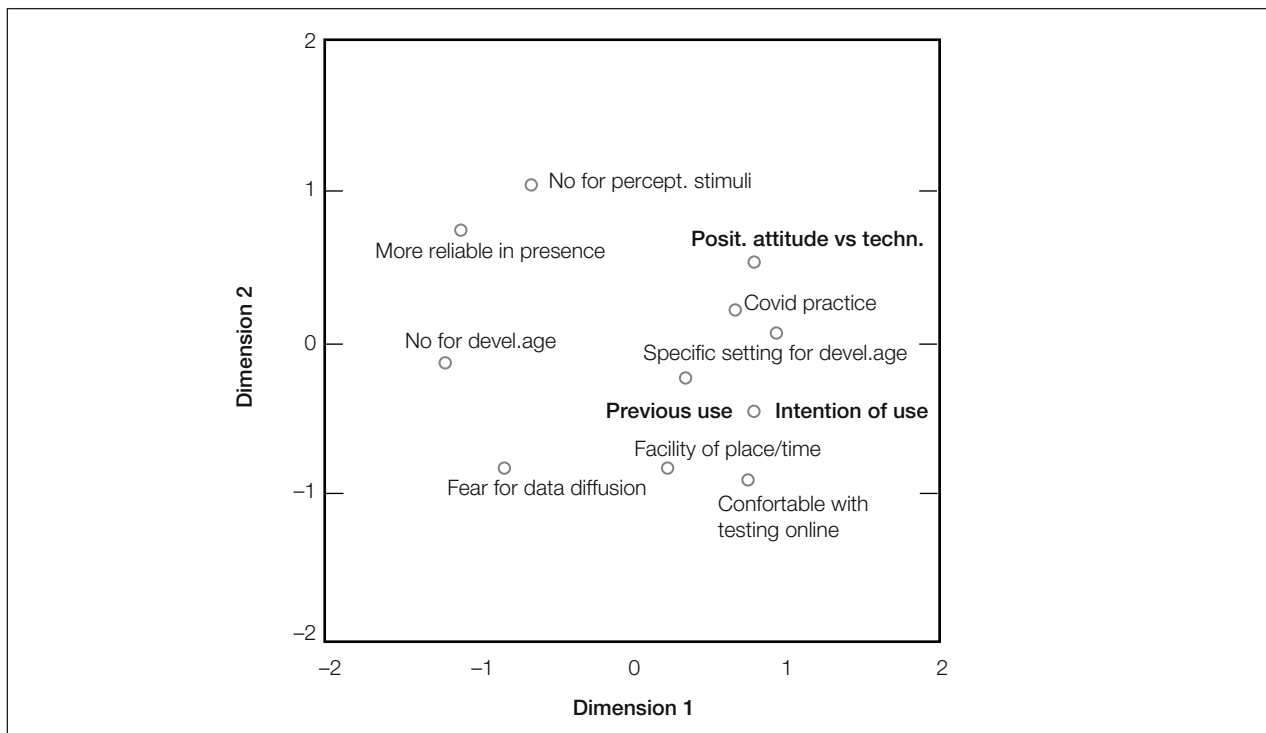
Note. Alienation of final configuration = .20; proportion of variance (RSQ) = .83.

CONCLUSION

The limitations of this preliminary survey, parallel to the EFPA's international study, were (a) the impossibility of reliably differencing subgroups of the sample, (b) the exploratory approach, and (c) the use of a general questionnaire about online assessment, without asking separately about the different instruments to use online (e.g., cognitive tests, personality inventories, development scales, verbal vs performance tests). A more detailed study should be carried (a) with a larger sample articulated in subgroups of psychologists, already specialized or during their training period, (b) with specific hypotheses about subgroups differences, and (c) with a questionnaire including specific sections on the cognitive, personality, verbal skills and performance variables to measure online. Furthermore, the tool could be improved and generalized administering it by means of different media, i.e. a computer, an app and a robotic assistant.

Despite the preliminary character of the study, the literature review on online testing, and the research data presented, allows some consideration regarding the psychologists' awareness of the difference between remote assessment, or through artificial agents, and the traditional method. The latter - although it can be supported by computerized formats, scoring or interpretations - is however based on the administration of a face-to-face interaction between the psychologist and the user, which long-distance or virtual relationships cannot reproduce in full.

From the results obtained in the study, it can be deduced that the overall attitude of the psychologists toward these innovative modalities of assessment is positive, depending also on the perception of usefulness deriving from the pandemic period. The intention to use online testing is very high, especially in the younger subgroup of the sample. The most predictive variables of the intention to use online testing are the absence of prejudicial refusal of a modality different from face-to-face, and the admission that online testing can also be useful for children and persons with frailty that need extra support or care, but with attention to specific settings for these particular participants. Moreover, the intention to use online testing is influenced by how comfortable the psychologist feels about this way of working, and by the practical convenience of using the test at the most appropriate times and places for the client.

The multidimensional analysis clearly shows that the attitude towards online testing, and the intention to use it, are located in the crossing dimensions of technical (psychometric) aspects and concrete usability.

In conclusion, the general consideration that can be deduced is that psychological testing in telematic mode cannot be considered a simple surrogate for traditional testing, as distance learning or smart working are in educational or work environments. The administration of tests electronically or through artificial agents requires adaptation studies and in many cases a reformulation of the tools that are offered online, taking into account the necessary formal changes as well as the standardization and calibration of the answers obtained.

Finally, for these innovative methods to be a useful advance in psychological assessment, even beyond the pandemic period, the need to prepare specifically validated tools must be underlined, through appropriate psychometric research.

There is also a need to raise awareness amongst psychologists and train those who intend to use it. This training should take place at various levels of university teaching, from basic courses to masters and specialisations; at the same time, professional psychologists could be sensitized to the correct use of the innovative techniques, and to accept them for what they can actually offer.

References

BAIG, M.Z. & KAVAKLI, M. (2019). A survey on psycho-physiological analysis & measurement methods in multimodal systems. *Multimodal Technologies and Interaction*, *3* (2), 37.

BREARLY, T.W., SHURA, R.D., MARTINDALE, S.L., LAZOWSKI, R.A., LUXTON, D.D., SHENAL, B.V. & ROWLAND, J.A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, *27* (2), 174-186.

BUCHANAN, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, *33* (2), 148.

BUKIE, O.F. (2014). Understanding technologies for e-assessment: A systematic review approach. *Journal of Emerging Trends in Computing and Information Sciences*, *5* (12), 936-947.

CONTI, D., COMMODARI, E. & BUONO, S. (2017). Personality factors and acceptability of socially assistive robotics in teachers with and without specialized training for children with disability. *Life Span and Disability*, *20* (2), 251-272.

CONTI, D., DI NUOVO, S., BUONO, S. & DI NUOVO, A. (2017). Robots in education and care of children with developmental disabilities: A study on acceptance by experienced and future professionals. *International Journal of Social Robotics*, *9*, 51-62.

CORTELLESSA, G., FRACASSO, F., SORRENTINO, A., ORLANDINI, A., BERNARDI, G., CORACI, L., DE BENEDICTIS, R. & CESTA, A. (2018). ROBIN, a telepresence robot to support older users monitoring and social inclusion: Development and evaluation. *Telemedicine and E-Health*, *24* (2), 145-154.

CULLUM, C.M., HYNAN, L.S., GROSCH, M., PARIKH, M. & WEINER, M.F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society: JINS*, *20* (10), 1028.

DE GRAAF, M.M.A., BEN ALLOUCH, S. & VAN DIJK, J.A.G.M. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human-Computer Interaction*, *34* (2), 115-173.

DI NUOVO, A., VARRASI, S., LUCAS, A., CONTI, D., McNAMARA, J. & SORANZO, A. (2019). Assessment of cognitive skills via human-robot interaction and cloud computing. *Journal of Bionic Engineering, 16*, 526-539.

DI NUOVO, S.F. & NARZISI, V. (2021). Telepsychology: Acceptability of online assessment and treatment in users and professional psychologists. *Mediterranean Journal of Clinical Psychology*, *9* (3).

FEINGOLD POLAK, R., ELISHAY, A., SHAHAR, Y., STEIN, M., EDAN, Y. & LEVY-TZEDEK, S. (2018). Differences between young and old users when interacting with a humanoid robot: A qualitative usability study. *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018*, Chicago, IL, USA, 107-108.

FISKE, A., HENNINGSEN, P. & BUYX, A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, *21* (5), e13216.

GRAZIOLI, S., MAURI, M., ROSI, E., VILLA, F., TIZZONI, F., TARABELLONI, A., TRABATTONI, S., MAURI, V., COLOMBO, P. & MOLTENI, M. (2022). Use of machine learning on clinical questionnaires data to support the diagnostic classification of attention deficit hyperactivity disorder: A personalized medicine approach. *European Psychiatry*, *65* (S1), 165-166.

IVANOVIĆ, M. & JAIN, L.C. (2013). *E-learning paradigms and applications: Agent-based Approach (Vol. 528)*. Springer.

OSGOOD, C.E., SUCI, G.J. & TANNENBAUM, P.H. (1957). *The measurement of meaning (Issue 47)*. University of Illinois Press.

PRESTON, J.A. & SHACKELFORD, R. (1999). Improving on-line assessment: An investigation of existing marking methodologies. *Proceedings of the 4th Annual SIGCSE/SIGCUE ITiCSE Conference on Innovation and Technology in Computer Science Education*, 29-32.

RICHES, S., AZEVEDO, L., VORA, A., KALEVA, I., TAYLOR, L., GUAN, P., JEYARAJAGURU, P., McINTOSH, H., PETROU, C. & PISANI, S. (2022). Therapeutic engagement in robot-assisted psychological interventions: A systematic review. *Clinical Psychology & Psychotherapy*, *29* (3), 857-873.

ROSSI, S., CONTI, D., GARRAMONE, F., SANTANGELO, G., STAFFA, M., VARRASI, S. & DI NUOVO, A. (2020). The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. *Robotics*, *9* (2), 39.

ROSSI, S., SANTANGELO, G., STAFFA, M., VARRASI, S., CONTI, D. & DI NUOVO, A. (2018). Psychometric evaluation supported by a social robot: Personality factors and technology acceptance. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 802-807.

SCASSELLATI, B., ADMONI, H. & MATARIĆ, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, *14*, 275-294. doi.org/10.1146/annurev-bioeng-071811-150036

SORA, B., NIETO, R., DEL CAMPO, A.M. & ARMAYONES, M. (2021). Acceptance and use of telepsychology from the clients'

perspective: Questionnaire study to document perceived advantages and barriers. *JMIR Mental Health*, *8* (10), e22199.

TAKAHASHI, H., BAN, M. & ASADA, M. (2016). Semantic differential scale method can reveal multi-dimensional aspects of mind perception. *Frontiers in Psychology*, *7*, 1717.

VALLEJO, M.A., JORDÁN, C.M., DÍAZ, M.I., COMECHE, M.I. & ORTEGA, J. (2007). Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire – 28 (GHQ-28) and the Symptoms Check-List – 90 – Revised (SCL-90-R). *Journal of Medical Internet Research*, *9* (1), e540.

VARRASI, S., DI NUOVO, S., CONTI, D. & DI NUOVO, A. (2018a). A social robot for cognitive assessment. *HRI'18 Companion: Conference on ACM/IEEE International Conference on Human-Robot Interaction, March 5-8, 2018, Chicago, IL*, USA, 269-270.

VARRASI, S., DI NUOVO, S., CONTI, D. & DI NUOVO, A. (2018b). Social robots as psychometric tools for cognitive assessment: A pilot test. *Springer Proceedings in Advanced Robotics - 10th International Workshop in Human-Friendly Robotics*, 99-112.

VARRASI, S., LUCAS, A., SORANZO, A., McNAMARA, J. & DI NUOVO, A. (2019). IBM cloud services enhance automatic cognitive assessment via human-robot interaction. In G. Carbone, M. Ceccarelli & D. Pisla (Eds.), *New trends in medical and service robotics*. Springer International Publishing.

VENKATESH, V., MORRIS, M., DAVIS, G. & DAVIS, F. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*, 425-478.

VIRKUS, S., LEOSTE, J., MARMOR, K., KASUK, T. & TALISAINEN, A. (2023, ahead-of-print). Telepresence robots from the perspective of psychology and educational sciences. *Information and Learning Sciences*.

WADSWORTH, H.E., DHIMA, K., WOMACK, K.B., HART JR.J., WEINER, M.F., HYNAN, L.S. & CULLUM, C.M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology*, *33* (8), 1040-1045.

WAINER, H., DORANS, N.J., FLAUGHER, R., GREEN, B.F. & MISLEVY, R.J. (2000). *Computerized adaptive testing: A primer*. Routledge.

ZHANG, G. & HANSEN, J.P. (2022). Telepresence robots for people with special needs: A systematic review. *International Journal of Human-Computer Interaction*, *38* (17), 1651-1667.

APPENDIX

Questionnaire items

	English version	**Italian version**
1	*Online testing can be as reliable as traditional testing*	Il questionario online può essere attendibile quanto quelli compilati in modo tradizionale
2	*Online testing is possible at the place and time most convenient for the user*	Chi risponde al questionario online può farlo nel luogo e nel momento della giornata che ritiene più opportuno
3	*I feel comfortable administering an online questionnaire*	Mi sento a mio agio nel far compilare un questionario online
4®	*The answered face to face questionnaire is more reliable than online*	Il questionario compilato in presenza dà risultati più affidabili di quello online
5®	*Some respondents could fear of making mistakes that cannot be corrected*	Qualcuno potrebbe esitare a rispondere alle domande online per paura di commettere errori che non può correggere
6	*Users think they get some health benefits from online testing*	Gli utenti pensano di poter ottenere dei benefici per la loro salute dopo la compilazione del questionario online
7®	*Online answers could be more accurate if the client can ask for explanations*	Se avessero la possibilità di chiedere spiegazioni sul contenuto di alcune domande, le risposte online degli utenti sarebbero più accurate
8	*The online questions are clear and easily understandable*	Le domande online sono scritte in modo chiaro e facilmente comprensibile
9	*I have the necessary technological resources to administer online testing (internet connection, suitable device)*	Ho le risorse necessarie per somministrare questionari online (connessione internet, dispositivo adeguato)
10®	*If the choice is possible, face to face testing is preferable*	Se avessi la possibilità di scegliere, preferirei far compilare il questionario faccia a faccia
11®	*Some respondent could fear that one's data and answers could be disclosed on the Internet*	Chi risponde online può avere paura che i propri dati e le proprie risposte possano essere diffusi su Internet
12	*The normative data of online tests are similar to those administered in presence*	I dati normativi dei test online sono analoghi a quelli della somministrazione in presenza
13	*Online testing requires specific preparation for the psychologist*	L'utilizzo dei test online richiede una specifica preparazione per lo psicologo
14	*Online testing requires specific preparation for the client*	L'utilizzo dei test online richiede una specifica preparazione per il cliente
15®	*Online testing is not suitable for the most clients with frailty*	Per i clienti più fragili è meglio non procedere alla somministrazione di test online
16	*For frail patients, the online test requires an adequate setting*	Per i clienti più fragili si può procedere alla somministrazione del test online se viene introdotto un adeguato setting
17	*All tests can be administered online*	Tutti i test possono essere utilizzati online

	English version	**Italian version**
18®	*Online testing is not suitable if perceptually complex materials are required*	Test che richiedono l'utilizzo di materiali percettivamente complessi non possono essere utilizzati online
19®	*Online testing is not suitable for clients in developmental age*	Per i clienti in età evolutiva è meglio non procedere alla somministrazione di test online
20	*For clients in the developmental age, an adequate online setting is necessary*	Per i clienti in età evolutiva si può procedere alla somministrazione del test online se viene introdotto un adeguato setting
21	*Covid-19 has enabled new opportunities for online testing*	Il Covid-19 ha fornito l'opportunità di usare nuovi processi digitali quali il testing online

Legenda. ® = reversed scoring.